ASSESSMENT OF GENERAL CAPABILITIES

SKILLS FOR THE 21ST-CENTURY LEARNER

FINAL REPORT

Claire Scoular, Dara Ramalingam,
Daniel Duckworth and Jonathan Heard

CARI

ACER

The Australian Council for Educational Research Ltd
19 Prospect Hill Road
Camberwell VIC 3124
Phone: (03) 9277 5555
ABN 19 004 398 145

www.acer.org

ISBN 978-1-74286-599-7

Recommended citation

Scoular, C., Ramalingam, D., Duckworth, D., & Heard, J. (2020). *Assessment of general capabilities: Skills for the 21st-century learner. Final report*. Australian Council for Educational Research. https://research.acer.edu.au/ar_misc/47

CONTENTS

LIST OF FIGURES

LIST OF FIGURES

HIGHLIGHTS

Addressing key issues

To understand development

Skill development frameworks are detailed descriptions of skills defined by their associated strands and aspects

To monitor growth

Skill development levels provide a framework in which to monitor growth and identify different levels of quality

To ensure alignment

An aligned approach can be managed through embedding of the aspects in curriculum, assessment *and* pedagogy

Generating better understanding of the skills

The project presents a valid and reliable assessment approach for measuring critical thinking, creative thinking and collaborative skills.

Problem-based learning modules are one effective way to measure multiple skills in different grades and learning areas.

The conceptualisation of the skills and their respective levels in the *skill development frameworks* were validated and informed by the assessment data

The assessment data indicates three separate constructs were visible and these constructs have relationships with one another. Of the three skills, critical thinking and creative thinking have the larger overlap conceptually

Assessments were well targeted to the selected grades (5 and 8). Only a few items were too easy or difficult for the students tested. It was possible for younger students to engage with more sophisticated problem-based scenarios if they were given the appropriate scaffolding, and the scenarios were framed appropriately

The embedment of the aspects in the *skill development frameworks* can be achieved at different grade levels, as evidenced by the majority of items across all three skills showing no difference in the way students of the same ability performed depending on their grade level

Critical thinking

- Applying logic was the most difficult aspect of critical thinking
- Identifying propositions consistent with another appeared a crucial first developmental step towards learning how to identify a logical conclusion of a proposition

Collaboration

- Communicating in collaborative groups was a common behaviour and appeared easy to do
- Regulating own contributions was generally difficult
- Resolving differences when deciding roles was significantly more difficult than resolving differences when deciding on the best ideas, suggesting the most difficult aspect of collaboration in the assessment is the employment of diplomatic strategies when members of the group are personally invested in some outcome contrary to your own

Creative thinking

- Coming up with multiple ideas was generally easy
- Generating original or novel ideas was more difficult

BACKGROUND

The Centre for Assessment Reform and Innovation (CARI) at the Australian Council for Educational Research (ACER) undertook a three year project to address the growing demands in the area of skill development.

There is increasing recognition that general capabilities, or 21st-century skills[1] as they are often called, are important for learning. There is growing consensus that these skills need to be cultivated to help learners succeed in a modern society based on knowledge and innovation and that embedding them within existing teaching practices should be a priority (World Economic Forum, 2016; Roseth et al., 2016). Broader ranges of skills, beyond literacy and numeracy, are increasingly visible and evident in national education policies and curricula. For example, the United Nations 2030 Agenda for Sustainable Development, and Sustainable Development Goal 4 'achieving an inclusive and quality education for all' has enhanced the focus on development of broader skills. Further, research indicates that cultivation of such skills within learning areas enhances student knowledge and application (Baghaei et al., 2007; Soller, 2001; Webb et al., 1998; Zhang, 1998).

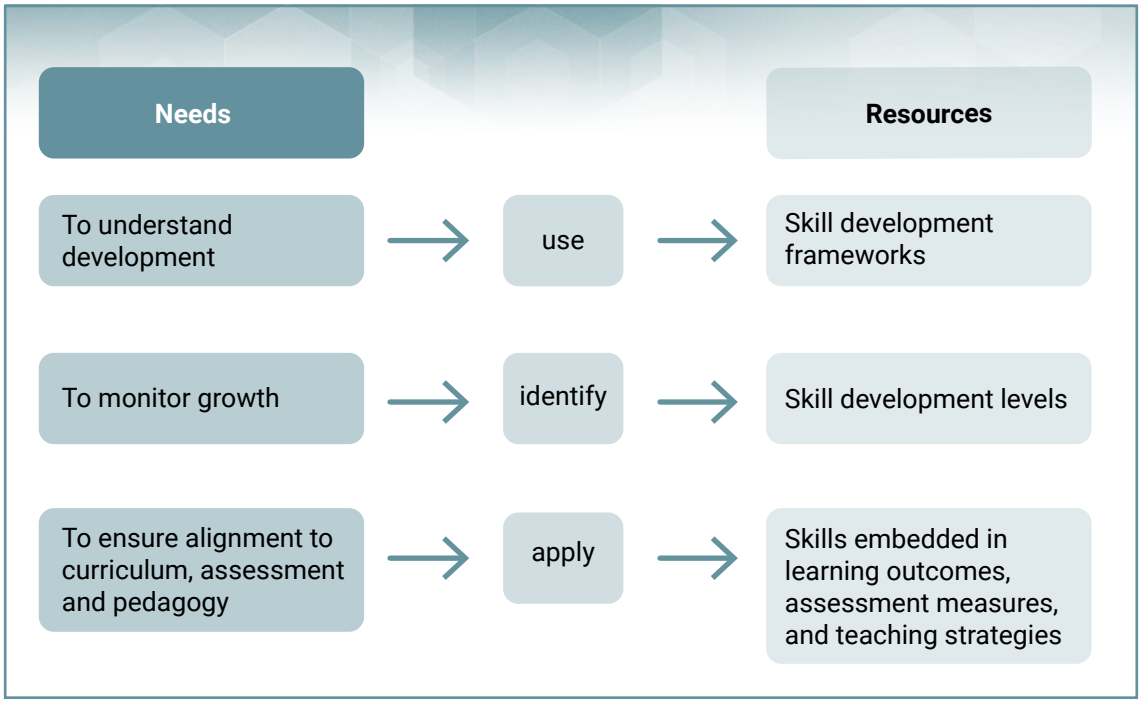
The focus of this project has been to develop an approach for teaching and assessing skills in the classroom. While there is wide recognition that students need to be better equipped with appropriate social and cognitive skills there is a lack of training to better equip educators to do this. The purpose of this project was to engage with educators to develop, trial, and validate resources at classroom level and ultimately equip them with the skills and resources they need to embed skills into their practice. Skill development frameworks are at the centre of these resources and support understanding of the skills. Over the life of the project, a research community has been established, workshops and masterclasses have been delivered to support professional development, and case studies have been captured.

The approach will be outlined in Part One of this report, and Part Two will outline how the approach is applied in the context of assessment.

[1] The term 21st-century skills is used to refer to a broad range of skills that are also sometimes referred to as general capabilities, transversal competencies, soft skills, or similar.

PART ONE
OVERVIEW OF ACER'S APPROACH TO SKILL DEVELOPMENT

ACER's approach to skill development has been underpinned by the identification of three evident needs: to understand development, to monitor growth, and to ensure alignment across curriculum, assessment, and pedagogy. To address these needs, the general capabilities project developed resources for each need and shown in Box 1. Through a combination of skill development frameworks, levels of skill development, and curriculum-orientated assessment and teaching tools, the project aims to equip teachers to measure and monitor the skills in their classroom, and better develop these skills in their students.



Box 1 Resources developed to meet the needs for skill development

1 USING SKILL DEVELOPMENT FRAMEWORKS TO UNDERSTAND DEVELOPMENT

At the centre of ACER's approach to assessing and teaching skills in the classroom is the need to understand the skills, and how skill development can best be supported. ACER has been investigating several skills in recent years and identified a subset as a focus of the general capabilities project: critical thinking, creative thinking and collaboration. ACER's selection was based on feedback from teachers in related projects as to which were most important or most familiar. Global studies were also reviewed to identify cross-country priorities.[2] The project approach is designed to be scaled up and is readily applicable to other skills.

In the 21st century, proficiency in critical thinking, creative thinking, and collaboration is highly valued within educational and professional settings. The degree to which they are defined, taught and assessed, however, is not well documented. While frameworks and definitions for skills are commonplace, they often lack sufficient detail for educators to understand how these skills manifest as observable behaviours in the classroom, and there is a distinct lack of evidence-based research. Many schools or systems are attempting to incorporate the assessment and teaching of skills, but most educators do not receive training to enhance understanding of the skills and how they can best be embedded into existing practices. This complicates efforts to develop these skills in students and to devise appropriate intervention strategies and assessment tools.

To assist in these efforts, the general capabilities project devised skill development frameworks for critical thinking, creative thinking and collaboration. These frameworks are designed to support researchers and educators to provide a clear model from which to base their understanding of the skills, and address the challenges associated with teaching and assessing them. ACER's detailed descriptions of the skills are a valuable resource in establishing consistent terminology in the wider community.

The skill development frameworks have a range of purposes:

- to provide detailed descriptions of skills as per their associated strands and aspects
- to identify levels of skill development
- to situate each skill within an education context and ensure it is goal orientated
- to provide a model for adopting consistent terminology
- to ensure that each skill is aligned in curriculum, assessment and pedagogy.

The skill development frameworks outline the aspects students engage in when they are applying the skills. The frameworks describe each skill within strands (core elements) that are then further delineated as aspects (sub-elements). Specifically, a strand refers to the overarching conceptual category for framing the skill, while the aspects refer to

[2] For example, one NEQMAP study reviewed policy, curriculum, and assessment documentation in 152 countries and identified that almost half those countries prioritised creative thinking, critical thinking, communication, and problem-solving skills (Roth et al., 2017).

specific behaviours associated with a strand. The aspects contained within the skill frameworks are designed to provide foci for teaching and form the basis of assessment. For example, educators can write assessment items to measure the specific aspects, or integrate teaching of an aspect into a lesson, rather than the whole skill.

The skill development frameworks characterise the skills as processes that are ultimately goal directed and purpose driven, whether that purpose is to solve a problem, complete a task, or decide on a course of action. In other words, the skill definitions are situated on the premise that there is purpose and necessity to employing the skill.

The definitions are intended to support understanding of the skill, not just for educators, but for students and the wider community. If a student can understand what the skills are and how they can be applied, particularly in enhancing knowledge in learning areas, then the skills are more accessible and applicable. With sufficient understanding of the skills, students can build a metacognitive understanding of their own development. They can understand what it means to be a good critical thinker, creative thinker or collaborator and reflect on their own ability to apply the skills in any given learning area.

Summaries of the skill development frameworks are presented in Figure 1.1 and in Appendix A. The full skill development frameworks include detailed definition descriptions of strands and aspects, supporting literature and research, and skill development levels.[3]

1.1 The applicability of the frameworks across learning areas

The ACER skill development frameworks convey that each skill consists of aspects that are described in general terms across learning areas. These aspects provide consistent terminology, and can be used to plan learning, or write and map assessment items. However, these general aspects – of critical thinking, creative thinking, and collaboration – may manifest themselves differently across learning areas. For students to develop in these aspects, the teaching and assessment of them needs to be translated into or embedded within the existing methodologies, conventions and 'ways of knowing' of each of the learning areas. This will give their application context to ensure they are relevant and can be sustainably integrated into lesson plans and learning outcomes across learning areas.

Further, even though the skill aspects can be described in a general way, this does not assume transferability. The ACER skill development frameworks address basic concepts about what it broadly means to think critically or creatively, or to collaborate, that can be generally applied across a wide range of learning areas and contexts. Defining the skills in general terms suggests that the skills can be described in a consistent way across learning areas, and that the relative differences in what makes someone more or less proficient in a particular skill is the same. This does not mean proficiency is transferable: demonstrating an aspect of a skill in one learning area does not mean that a student can do it in another. Discipline- or context-specific knowledge is integral to being able to demonstrate a skill within that context, and as proficiency of skill increases in one particular learning area, that same level of skill will not necessarily transfer to other contexts where deep knowledge is absent.

[2] The skill development frameworks are publicly available. Critical thinking: https://research.acer.edu.au/ar_misc/41/; Creative thinking: https://research.acer.edu.au/ar_misc/40/ ; Collaboration: https://research.acer.edu.au/ar_misc/42/

Critical thinking

Strand 1 Knowledge construction

Aspect 1.1 Identifies gaps in knowledge

Aspect 1.2 Discriminates amongst information

Aspect 1.3 Identifies patterns and makes connections

Strand 2 Evaluating reasoning

Aspect 2.1 Applies logic

Aspect 2.2 Identifies assumptions and motivations

Aspect 2.3 Justifies arguments

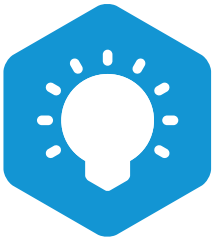
Strand 3 Decision-making

Aspect 3.1 Identifies criteria for decision-making

Aspect 3.2 Evaluates options

Aspect 3.3 Tests and monitors implementation

Creative thinking



Strand 1 Generation of ideas

Aspect 1.1 Number of ideas

Aspect 1.2 Range of ideas

Strand 2 Experimentation

Aspect 2.1 Shifting perspectives

Aspect 2.2 Manipulating ideas

Strand 3 Quality of ideas

Aspect 3.1 Fitness for purpose

Aspect 3.2 Novelty

Aspect 3.3 Elaboration

Collaboration

Strand 1 Building shared understanding

Aspect 1.1 Communicates with others

Aspect 1.2 Pools resources and information

Aspect 1.3 Negotiates roles and responsibilities

Strand 2 Collectively contributing

Aspect 2.1 Participates in the group

Aspect 2.2 Recognises contributions of others

Aspect 2.3 Engages with roles and responsibilities

Strand 3 Regulating

Aspect 3.1 Ensures own contributions are constructive

Aspect 3.2 Resolves differences

Aspect 3.3 Maintains shared understanding

Aspect 3.4 Adapts behaviour and contributions for others

Figure 1.1 Summary of ACER's skill development frameworks for critical thinking, creative thinking and collaboration

1.2 An evidence-based approach

One of the hurdles in teaching and assessing skills is that there is an abundance of surface-level descriptions. With so many general definitions of critical thinking, creative thinking, and collaboration, educators are unclear as to which definition to adopt. Further, these surface-level definitions don't allow for deep understanding of the skills in practice, and therefore it is difficult to meaningfully integrate them into teaching practices.

ACER' approach has been to use evidence to create the skill development frameworks. While their conceptualisation and associated levels of skill development were initially hypothesised, measures were taken over the last several years to increase their evidence base and validity through:

- synthesising the literature
- reviewing existing research
- integrating educators' classroom terminology
- mapping existing assessment data to the aspects and levels
- using the aspects to develop new formative assessments and observation tools
- applying the skill development frameworks in multiple national and international projects
- embedding aspects into lesson plans and teaching modules
- comparing and mapping aspects to national curricula documentation.

Together, these measures have ensured the skill development frameworks and the associated levels of skill development are validated resources that are fit for purpose.

2 IDENTIFYING SKILL DEVELOPMENT LEVELS TO MONITOR GROWTH

ACER's perspective of skill development is centred on and emphasises the notion of growth. Skills can be defined with a growth perspective, can be improved through teaching and intervention, and can be measured and monitored.

For each of the respective skills, levels of skill development are used to describe how growth in a particular skill can be demonstrated, and how students move from early, to more advanced application and understandings. These levels of skill development focus on assessing and monitoring growth over time, and are underpinned by an understanding that students of the same age and in the same year of school can be at very different points in their learning and development. Therefore, they are not linked to specific years of schooling. When assessments provide information about where students are in their progress at the time of assessment, they also provide a basis for monitoring individual progress over time. Assessments of progress are an alternative to judging success only in terms of year-level standards.

While progress can be described in a general way – what a highly proficient critical thinker demonstrates compared to a less proficient critical thinker, for example – the application of the skill by an individual still depends on the specific learning area.

Levels of skill development can support understanding of the skills and how they develop. They can also support teachers to identify gaps in a learning area, where some students may require further assistance.

Table 2.1 presents an excerpt from the levels of skill development for collaboration. The full skill development levels for critical thinking, creative thinking, and collaboration are presented in the respective framework documents.[4] The levels are divided into strands, and labelling of aspects allows educators to identify and monitor student progress within and across each of the aspects. To ensure an evidence-based approach, these levels have been, and continue to be, validated and corroborated with assessment data.

Skill level	Building shared understanding	Collectively contributing	Regulating
Medium	Learners ask for justification of responses or perspective provided. (Aspect 1.1)	Learners acknowledge that others may have a different perspective, and that based on these perspectives, others' contributions may be beneficial to the group as a whole. They understand and incorporate the contributions of others into their own work. (Aspect 2.2)	Learners identify own strengths and weaknesses in relation to the progress of the group task as whole. (Aspect 3.1) Learners make constructive but unsuccessful attempts to resolve differences. (Aspect 3.2) Learners act to maintain shared understanding such as by reiterating or finalising goals, strategy, and roles in more complex tasks. (Aspect 3.3) Learners require feedback from others or explicit requests before they modify or tailor their communication style or behaviour. (Aspect 3.4)
Low–Mid	Learners ask questions or for clarification from others. They will communicate about the related task and respond to contributions of others. (Aspect 1.1) Learners identify that they may not have all of the information required and pool some resources and information with others. (Aspect 1.2) Learners negotiate roles but without considering the expertise, information, or skills held by other group members. (Aspect 1.3)	Learners participate in all necessary tasks throughout the task. Learners maintain a single strategy throughout. Learners collaborate successfully to achieve a straightforward goal. (Aspect 2.1) Learners understand that others may have an alternative perspective. They listen to and acknowledge the perspective of others. (Aspect 2.2) Learners show a willingness and readiness to be involved in the group. They take responsibility for some of the actions determined by their role and provide feedback on their individual tasks. (Aspect 2.3)	Learners reflect on the quality and relevance of their own contributions. (Aspect 3.1) Learners discuss differences of opinion or perspective with others and give careful consideration of the views of others. They comment on differences, but are often unable to resolve them. (Aspect 3.2) Learners act to maintain shared understanding through reiterating goals, strategy, and roles in basic tasks. (Aspect 3.3)

Table 2.1 Excerpt from the levels of skill development for collaboration (Scoular et al., 2020)

[4] The skill development frameworks are publicly available. Critical thinking: https://research.acer.edu.au/ar_misc/41/; Creative thinking: https://research.acer.edu.au/ar_misc/40/ ; Collaboration: https://research.acer.edu.au/ar_misc/42/

3 ENSURING ALIGNMENT BY EMBEDDING SKILLS IN CURRICULUM, ASSESSMENT AND PEDAGOGY

Curriculum, assessment and pedagogy are strongly linked (see Figure 3.1). What the curricula defines and sets out influences what is taught, and it follows, what is assessed. While the links between curriculum, assessment and pedagogy tend to be well-established in learning areas with a long history, this is less the case with skills. Embedding skills requires educators to have expertise in curriculum development, assessment design and pedagogical strategies in order to achieve the support structures and substance required to form a coherent approach. So far, many education systems have tried to embed skills from a curricular base, with a few initiating assessment reform (Care et al., 2016).

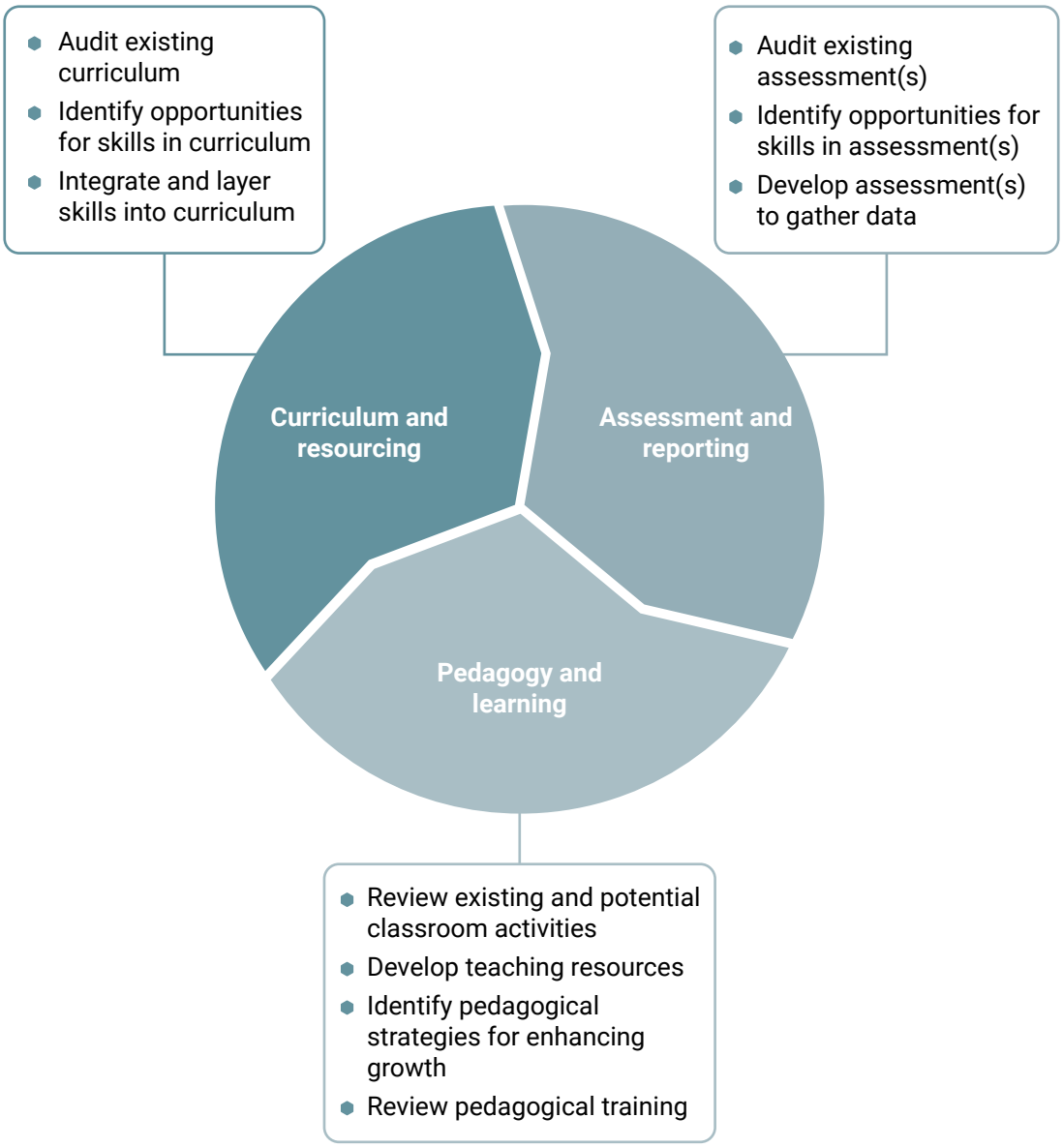


Figure 3.1 ACER's alignment model for embedding skills in education systems

In a fully integrated approach, learning outcomes comprise both skill development and learning area knowledge. The curriculum sets out the learning outcomes and provides a framework to identify where and what skills are presented in different learning areas. Assessment of skills provides an understanding of where students are at in their development, how they progress over time, and whether they achieve those specific learning outcomes. Pedagogical strategies provide the means for skills to be taught within existing learning areas to develop those learning outcomes.

As shown in Figure 3.1, for education systems, an aligned approach to embedding skills involves many complex and iterative activities. However, the aspects outlined in ACER's skill development frameworks provide a basis on which to proceed. At the core of an aligned approach is an understanding of what needs to be integrated. This can be provided when educators have sufficient understanding of what the skills are, how they can be applied, and how they develop in students.

An aligned approach can also be achieved at school and classroom level. The skills associated with the aspects can be mapped to learning outcomes in curriculum, educators can design classroom-based assessments that elicit behaviours associated with aspects, and teaching strategies can be used in which those aspects can be developed.

Table 3.1 presents an example of how this can applied in practice at school and classroom level. The aspects provide a more manageable and sustainable approach to embedding skills into existing practice. Instead of attempting to teach or assess a skill in its entirety, educators can focus first on the aspects with a view to covering a wider representation of the skill as their understanding builds. In the specific examples in Table 3.1, assessment criteria are taken from the skill development levels in the ACER frameworks, and learning outcomes are taken from the Australian curriculum.

Table 3.1 Examples of aspect application within curriculum, assessment, and pedagogy

Skill	Aspect	Curriculum	Assessment	Pedagogy
		Learning outcome	**Assessment criteria**	**Teaching strategy**
Collaboration	**Aspect 2.1 Participates in the group**	Personal and Social Capability **Element:** Social management **Sub-element:** Work collaboratively: Contribute to groups and teams, suggesting improvements in methods used for group investigations and projects	**High:** Learners participate throughout the task and try alternative strategies or multiple attempts during difficult tasks. **Mid:** Learners participate in all necessary activities in simple tasks maintaining a single strategy throughout. **Low:** Learners take action in the task but don't reach the end of the task.	Provide each learner with a different set of resources, leading to a task that requires all learners' input into the group's shared task.
	Aspect 3.2 Resolves differences	Personal and Social Capability **Element:** Social management **Sub-element:** Negotiate and resolve conflict: Identifies causes and effects of conflict, and practises different strategies to diffuse or resolve conflict situations	**High:** Learners resolve differences, explaining and justifying their understanding, leading to optimal collaboration. **Mid:** Learners can identify the cause and effect of conflicts and make constructive attempts to resolve differences by negotiating, debating and arguing their views. **Low:** Learners discuss differences of opinion or perspective with others and give careful consideration of the views of others. They comment on differences, but are often unable to resolve them.	Provide an authentic problem-based learning task that creates differing perspectives and solution ideas.
Critical thinking	**Aspect 1.2 Discriminates amongst information**	Critical and creative thinking **Element:** Inquiring – identifying, exploring and organising information and ideas **Sub-element:** Organise and process information: critically analyse information and evidence according to criteria such as validity and relevance	**High:** Learners can distinguish factual information from opinions and assertions, while recognising the potential value of each. **Mid:** In familiar, constrained contexts, learners can distinguish more reliable from less reliable information using objective criteria that are about evaluating quality **Low:** Learners discriminate between information sources using subjective criteria such as familiarity, accessibility, or alignment with their own views.	Teach students to evaluate texts they encounter using the 'C.R.A.A.P test' (currency, reliability, authority, accuracy and [author's] purpose).
	Aspect 2.3 Justifies arguments	Critical and creative thinking **Element:** Analysing, synthesising and evaluating reasoning and procedures **Sub-element:** Evaluate procedures and outcomes: explain intentions and justify ideas, methods and courses of action, and account for expected and unexpected outcomes against criteria they have identified	**High:** Learners can construct cogent arguments for and against a proposition – or for competing propositions – with explanations, supporting evidence, rebuttal and counter rebuttal. They can use inference to develop multiple plausible interpretations. **Mid:** Learners develop structured arguments for or against a proposition with some reasons and explanation. They use inference to develop a plausible interpretation. They can reflect on and explain their reasoning for claims they make. **Low:** Learners construct simple arguments supported by subjective reasoning, or plausible reasoning, in familiar, concrete contexts. They tend to use induction from experience of the world rather than deduction from rules, conditions or premises, and reach naive conclusions. In more abstract contexts, they use circular logic to articulate an argument.	Probing – or Socratic – questioning to elicit students' reasoning, or using scaffolds for thinking (e.g. Claim, Support, Question routine) and argument formulation (e.g. TEEL, or CEER [claim, explain, evidence, relevance].

Table 3.1 Examples of aspect application within curriculum, assessment, and pedagogy (Continued)

Skill	Aspect	Curriculum	Assessment	Pedagogy
		Learning outcome	**Assessment criteria**	**Teaching strategy**
Creative thinking	**Aspect 2.1 Shifting perspective**	Critical and creative thinking **Element:** Generating ideas, possibilities and actions **Sub-element:** Consider alternatives: identify situations where current approaches do not work, challenge existing ideas and generate alternative solutions	**High:** Learners demonstrate a willingness to experiment, shifting beyond conventional perspectives leading to new possibilities. They question and renegotiate the boundaries of the task to navigate around possible constraints. They test out multiple pathways, even those that seem unlikely. **Mid:** Learners can shift perspective, thinking about the task in a different way and considering the task from a range of conventional perspectives. They are willing to test out an alternative pathway. **Low:** Learners view the task through their single perspective without consideration of what the task elements can be changed, or considering alternative perspectives or pathways.	Teach techniques for reconsidering problems and situations, such as a Circle of Viewpoints routine, Six Thinking Hats or a PO disruption.
	3.2 Novelty	Critical and creative thinking **Element:** Generating ideas, possibilities and actions **Sub-element:** Imagine possibilities and connect ideas: Combine ideas in a variety of ways and from a range of sources to create new possibilities	**High:** Learners develop some original ideas containing concepts less familiar to them beyond their social context. **Low:** Learners present ideas that are obvious or conventional and contain concepts that are already familiar to them.	Challenge students to consider a range of novel solutions or ideas by facilitating an extended brainstorm (i.e. beyond fluency of ideas) to encourage flexibility or range, originality and elaboration of thinking.

PART TWO
AN APPROACH TO ASSESSING GENERAL CAPABILITIES

Part Two of this report outlines the assessment component of the general capabilities project.

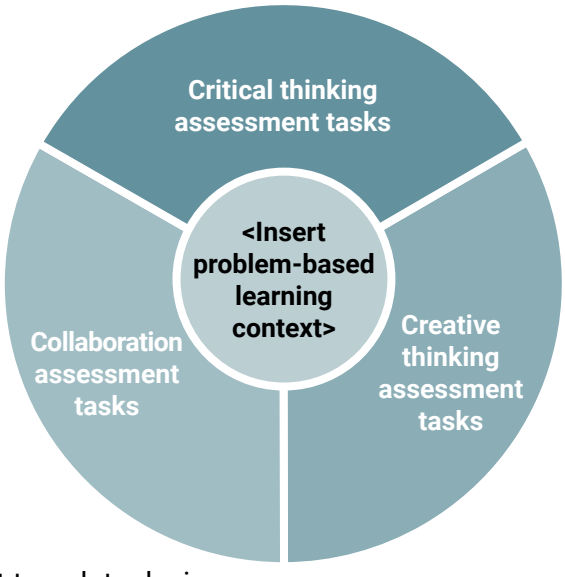
Skills are not only complex in nature, but also complex to assess. While it might be possible to measure a skill discretely using highly contrived and dissociated items, such assessments would likely fail to capture how these skills manifest when applied to the sorts of real-world, problem-solving contexts in which they are jointly required and are most often valued. To measure skills in authentic use therefore requires innovative methods of assessment. Further, in the context of classroom assessment, teachers need a reference framework to guide their judgements and observations of students. Teachers also need flexibility to adapt or develop assessments to a learning area, context, or environment.

The ACER general capabilities project developed an assessment template that encompasses built-in tasks and associated scoring systems, but is flexible enough to allow teachers to adapt the learning area or problem scenario to suit the grade level, or curriculum topic of current focus.

The assessment template:

- measures multiple skills
- is problem-based and authentic
- is domain orientated
- maps to skill development levels.

The design for the assessment template is presented in Box 2 and shows how the measurement of each of the skills is through authentic, problem-based learning tasks.



Box 2 Assessment template design

4 ASSESSMENT TEMPLATE

The assessment template is a consistent 'shell' for task design that provides a general approach across skills, grades, and learning areas. The template comprises 10 tasks that take from 5 minutes to 30 minutes each to complete and each focuses on measuring one of the three skills being assessed. The students primarily work in groups of three, although some tasks focusing on critical and creative thinking require individuals to work independently before returning to the group. Table 4.1 presents the assessment template with problem-solving stages, the associated assessment tasks, the task objective and the skill under investigation.

The assessment template is consistent regardless of the content. For example, it has the same set of tasks for a humanities-orientated assessment, as it does for a STEM (science, technology, engineering and mathematics) one. This provides an opportunity for teachers to 'plug and play', using the template to develop their own assessment tasks by embedding the content of their choice, within the template structure, and its associated measures. It can be adapted for an online or offline classroom-based setting.

Table 4.1 Assessment template

Problem-solving stage	Assessment task	Task objective	Central skill assessed
Problem presentation	Pre-learning/context and presentation of problem		n/a
Understanding the problem	**1** Analysing an FAQ	Understand the community's perception of the problem-solving scenario	Critical thinking
	2 Critical analysis of opinion pieces	Understand the assumptions and conclusions in opinions on the problem-solving scenario	Critical thinking
Initial idea generation	**3** Individual idea generation	Generate some initial ideas to support the problem-solving scenario	Creative thinking
	4a & **4b** 'Best' idea; evaluation of own and others' best ideas	Evaluate the strengths and weaknesses of each group member's idea	Critical thinking
	4c Selection of group's best idea	Work as a group to discuss and select the most creative idea to support the problem-solving scenario	Collaboration
	5 Reflection		n/a
Further investigation	**6** Role agreement	Use a group chat to negotiate a role in the group	Collaboration
	7 Reflection		n/a
	8a Individual research of the topic; potential adjustments to the group's best idea	Conduct role-specific research and use research results to improve your group's most creative idea	Critical thinking
Improvements and group decision	**8b** Sharing improvements and making a final group decision	Share information with the group to improve the group's idea	Collaboration
	9 Individual record of the group's agreed idea	Explain own understanding of the final idea agreed to by the whole group	Critical thinking
Communication of solution	**10** Explanation of the group's agreed best idea	As a group, record the group's final idea for submission to an entity related to the problem-solving scenario (e.g. a local council)	Creative thinking

4.1 Measuring multiple skills

Throughout the problem-solving process, the students need to employ a number of cognitive and social skills to define the problem, plan an approach and execute strategies in order to arrive at a solution to address the problem. When solving a complex problem in real-life, critical thinking skills are not used on their own, they are supported by the application of other social and cognitive skills such as collaboration and creative thinking to arrive at novel and workable solutions. Skills such as creative thinking and critical thinking often work in tandem and are likely to be highly interrelated (Sternberg, 1998). These interrelations need to be carefully addressed in the design of the assessment and scale development (Ercikan & Oliveri, 2016).

4.2 Problem-based and authentic

Problem-based learning (PBL) involves working through and reflecting on problems in small self-directed groups, with guidance from teachers as facilitators (Maudsley, 1999). In PBL, the context for learning is set via a real-world problem with multiple dimensions, around which a unit of work is planned (Parker & Thomsen, 2019).

According to Parker and Thomsen (2019), students undertaking PBL must:

- think critically about information and ideas
- think creatively about the problem
- work collaboratively to find the best solution.

Using PBL is a means of assessing the skills in authentic application, not in the abstract. The tasks devised in the ACER assessment are intended to have real-world relevance to stimulate motivation and engagement in students. The tasks in the ACER assessment are created along a PBL pathway, which assesses these skills as they are manifested, rather than inferring them retrospectively from the end-product.

There has been a focus in the literature on teaching the skills of problem-based or inquiry-based learning (Hmelo-Silver, 2004) – most likely because problem-solving is one of the most frequently mentioned 'in demand' skills and features consistently across skills frameworks. Complex problem-solving refers to 'the capacities to solve novel, ill-defined problems in complex, real-world settings' (World Economic Forum, 2016, p. 16). Complex problem-solving provides a rich, extended activity for students to use the range of skills that ACER is interested in measuring. Therefore, each of the skills presented in the ACER approach are contextualised in complex problem-solving activities. Although the problem tasks are primarily positioned as the context for students to work collaboratively with their classmates to come up with feasible solutions, the problem tasks are designed to give students the opportunity and time to engage and demonstrate the skills. By nature, 21st-century learning activities are often open-ended, involve unbounded sets of information, and there may be ongoing redefinition of the goal of the task (Scoular, 2019). It is important that students develop skills to establish and adapt goals according to available information, seek out relevant and valid information for the task, and continually monitor their own progress.

The ACER assessments adopt the PBL approach (adapted from Reeves et al., 2002) and offer the following features:

- have real-world relevance
- are ill-defined and require task definition
- are complex and need to be investigated
- allow for different perspectives and the use of a variety of resources
- provide the opportunity to collaborate
- provide the opportunity to reflect
- can be integrated across different subject areas
- can be seamlessly integrated with assessment
- create polished products valuable in their own right
- allow for competing solutions and diversity of outcomes.

The classroom provides a fertile ground to observe students demonstrating skills through real-life problem-solving scenarios or contexts. However, to measure skills in the classroom in a valid and reliable manner is very challenging. It can be challenging to identify 'good' problem tasks – tasks that are relevant and of interest to students and therefore motivate them to remain engaged over an extended period of time. The approach adopted in this project was to present authentic problem tasks to increase student interest and engagement, and support application of the skills to real-world contexts.

4.3 Domain orientated

The aspects in each of the ACER skill development frameworks need to be translated into or embedded within the learning areas. This gives the application context so as to ensure they are relevant and can be sustainably embedded into lesson plans and learning outcomes across learning areas.

A limitation of domain-specific assessment is that while skills do relate strongly to students' abilities to solve a routine task in a specific learning area, the inference is limited only to the domain in which the problem was framed (Greiff et al., 2014). The view that domain-specific problems only tell you about a student's ability within that domain is also shared by Care et al. (2015) in the Assessment and Teaching of 21st Century Skills study. The ACER approach addresses these tensions by presenting a domain-orientated rather than domain-specific option. This ensures that the assessment is still relevant to and contextualised within a learning area, yet it is not dependent upon extensive knowledge of that learning area. To this end, the ACER assessments can better target and measure the application of a skill, as distinct from the impact of knowledge of a learning area.

4.4 Mapping to skill development levels

The skill development frameworks provide a much clearer picture of what a skill 'is made of' and thus provide a narrower focus when developing assessment items. The assessment items are targeted to aspects, rather than the overarching definition

of the skill. For example, assessment items are not measuring critical thinking as a whole, but instead measuring the aspects of 'discriminates amongst information', or 'identifies criteria for decision-making', or 'justifies arguments'. The assessment tasks are designed to elicit specific aspects of the skills, leading to the identification of associated behaviours, and the recognition of different levels of skill development. The skill development levels both inform, and are informed by the assessment items.

5 ASSESSMENT TOOLS

For the purposes of the general capabilities project, two assessment modules were developed from the assessment template and targeted to Grade 5 and Grade 8 students. Developing assessments across different age groups was intended to provide evidence of the maturation of the skills and allow for student growth to be monitored across the grades. One assessment was orientated to a humanities context (refugee resettlement module), and the other to a STEM context (sculpture design module). Assessments were developed across different learning areas to compare skill development in different learning areas.

5.1 Assessment instrument overview

The ACER general capabilities assessment provides students (collaborating in groups of three) with a set of interrelated tasks unified around a problem-solving scenario. The scenarios are designed to be open-ended, require significant critical analysis and support a broad range of creative solutions. The test software differs from typical standardised tests that are typically organised by a test interface, which separates the test instructions, test navigation, stimulus content and answer space.

The assessment uses a combination of HTML pages and software products from Gsuite by Google to host and deliver the test content to students via an internet browser on a school computer. The task documents are filed in Google Documents. These documents contain task instructions and proformas into which students can enter information as the response format. A task typically comprises multiple assessment items. In addition, Google Forms is used to present multiple-choice items and reflection questionnaires for students to answer multiple-choice questions about their experience of working in a group. Google Hangouts is used to host chats between group members when students collaborate on a task. HTML pages are used to simulate multi-modal web-based information and resources that students can access to support some problem-solving activities. Google Drive is used to host the Google Documents and provides students with a way of navigating between tasks.

Prior to assessment administration, students are given access to some background learning content that provides an overview of the problem-solving scenario delivered by their classroom teacher. The background learning is provided to ensure that all participants have the same minimum level of knowledge about the problem-solving scenario.

5.2 Assessment modules

Humanities assessment: Refugee resettlement module

The refugee resettlement module's background learning introduces students to the concept of migration and the factors that contribute to refugees seeking to migrate from their home country. The problem-solving scenario involves resettling 120 refugees into an Australian suburb. The overarching problem-solving objective of the module is to develop a plan that would help resettle refugees within Australia and the local community.

STEM assessment: Sculpture design module

The sculpture design module's background learning introduces students to the concept of physical balance and provides students with content and formulae relating to centre of gravity. The problem-solving scenario presents students with a sculpture design competition hosted by a non-for-profit organisation. The competition challenges students to develop a sculpture design concept that utilises centre of gravity and which communicates an artistic message.

6 METHOD

6.1 Scoring assessment items

Most of the items in the assessment are constructed-response items. Scoring guides were developed and provide scoring criteria for every item. The scoring guides for each module are conceptually the same since each module was developed from the same assessment template, therefore the tasks and items are the same. However, the scoring guides across modules are adapted to match the problem-solving scenario and provide examples of student responses that are different across contexts and modules.

6.2 Mapping assessment items to the skill development frameworks

The items within tasks that comprise the assessment modules are based on the strands and subsequent aspects of the skill development frameworks. These skill definitions are central to the process of instrument development because they provide a theoretical underpinning for the assessment and a way of describing its content.

Table 6.1 shows the total number of items and score points attributed to each skill. Approximately 50 per cent of the items measure critical thinking; 30 per cent collaboration; and, 20 per cent creative thinking. The proportions of items by skill mirrors real-world problem-solving frameworks (e.g. design thinking) and typical workplace processes. Therefore, most items relate to critical thinking and are complemented by creative thinking and collaborative items.

Table 6.1 Number of items and score points by skill

Critical thinking			Creative thinking			Collaboration		
Strand/aspect	**Total (Items)**	**Max. total (Score points)**	**Strand/aspect**	**Total (Items)**	**Max. total (Score points)**	**Strand/aspect**	**Total (Items)**	**Max. total (Score points)**
Strand 1: Knowledge construction			**Strand 1: Generation of ideas**			**Strand 1: Building shared understanding**		
Aspect 1.1: Identifies gaps in knowledge	1	1	Aspect 1.1: Number of ideas	1	3	Aspect 1.1: Communicates with others	3	6
Aspect 1.2: Discriminates amongst information	9	9	Aspect 1.2: Range of ideas	1	3	Aspect 1.2: Pools resources & information	1	1
Aspect 1.3: Identifies patterns & makes connections	1	1				Aspect 1.3: Negotiates roles & responsibilities	1	1
Total (Strand 1)	**11**	**11**	**Total (Strand 1)**	**2**	**6**	**Total (Strand 1)**	**5**	**8**
Strand 2: Reasoning			**Strand 2: Experimentation**			**Strand 2: Collectively contributing**		
Aspect 2.1: Applies logic	3	3	Aspect 2.1: Shifting perspective	1	1	Aspect 2.1: Participates in the group	1	3
Aspect 2.2: Identifies assumptions & motivations	11	12	Aspect 2.2: Manipulating ideas	0	0	Aspect 2.2: Recognises contributions of others	2	2
Aspect 2.3: Justifies arguments	0	0				Aspect 2.3: Engages with role & responsibilities	0	0
Total (Strand 2)	**14**	**15**	**Total (Strand 2)**	**1**	**1**	**Total (Strand 2)**	**3**	**5**
Strand 3: Decision-making			**Strand 3: Quality of ideas**			**Strand 3: Regulating**		
Aspect 3.1: Identifies criteria for decision-making	0	0	Aspect 3.1: Fitness for purpose	2	3	Aspect 3.1: Ensures constructiveness of own contributions	3	3
Aspect 3.2: Evaluates options	7	13	Aspect 3.2: Novelty	2	2	Aspect 3.2: Resolves differences	4	4
Aspect 3.3: Tests & monitors implementation	0	0	Aspect 3.3: Elaboration	2	2	Aspect 3.3: Maintains shared understanding	1	1
						Aspect 3.4: Adapts behaviour & contributions for others	0	0
Total (Strand 3)	**7**	**13**	**Total (Strand 3)**	**6**	**7**	**Total (Strand 3)**	**8**	**8**

The assessment was not intended to cover all aspects of the skills, nor to cover each skill equally, but rather to ensure some coverage of most aspects that could be practically addressed in an authentic scenario administered in a single assessment. This means that some aspects are not covered in this assessment design.

6.3 Participants

The ACER general capabilities assessment trial was conducted around Australia. The trial did not employ random sampling methods, but instead used networks of schools interested in the trial. The participating schools were self-selected, volunteered to participate, and do not reflect a representative and unbiased population sample.

A summary of the number of schools and students that participated in the trial is shown in Table 6.2.

Table 6.2 Participants in the 2018/2019 trial

State	No. of schools	No. of students
Queensland	13	393
South Australia	5	331
Victoria	7	334
New South Wales	8	802
Total	**33**	**1860**

7 ANALYSIS

Psychometric and qualitative analyses were conducted in order to identify the robustness, reliability, and validity of the assessment in measuring the skills and to support greater understanding of the skills.

While both the humanities and STEM assessments were administered to Grades 5 and 8, there were insufficient cases of the STEM assessment required to run the Item Response Analysis. Therefore, the following results relate to the humanities assessment data only.

7.1 Item Response Theory

One of the objectives of this project was to explore whether the items that measure a specific skill work together to form a single variable that represents that skill. To do this, Item Response Theory (IRT) was applied. IRT investigates the items for consistency in relation to each other to support the construct validity of the assessment. If the items do not present in a cohesive manner they may need to be

adjusted or omitted to allow for clearer interpretation of the skill. Items omitted may have perfectly acceptable statistical qualities but if they deviate from the dimension under measurement then they may not be representative of the skill in question, and are therefore not relevant to the current assessment.

An IRT calibration was completed for each skill individually since each skill is identified as an individual construct. The fit of the items and the item discrimination were reviewed, as well as indices of reliability.

- The measure of fit indicates whether the items are assessing a unidimensional underlying construct. Therefore, in this project, the fit is interpreted as indicating whether the items in the assessment are each measuring either students' critical thinking, creative thinking, or collaboration.
- Item discrimination is used to assess item quality and identify item capacity to separate students who are at different levels of ability. To calculate the item discrimination index, a point-biserial correlation between each item and the overall test score can be used. This correlation describes the extent to which an item is consistent with other items.
- Item separation indices are good estimates of reliability in assessments. High item separation reliability is important for establishing that items are well separated along the continuum (Wright & Stone, 1979). Further, high item separation reliability provides sound evidence of construct validity (Wright & Masters, 1982).

Critical thinking

All 24 items for critical thinking were calibrated together. A summary of the output is presented in Table 7.1.[5] The item separation reliability is high at 0.987. This index indicates that items on the continuum were well separated and differentiated well between students' critical thinking abilities.

All of the items present a fit value within the confidence interval range. This suggests that the items represent a single construct, interpreted as critical thinking, and provide good item quality in relation to measuring that skill.

Item 17 indicates marginal discrimination as presented in Table 7.1. The Item Characteristic Curve shown in Figure 7.1 shows student ability increasing on the horizontal axis as the probability of being assigned the higher score category increases on the vertical axis, with low ability students on the left and higher on the right. The dotted line represents the observed curve of an actual response, and the solid line represents the expected, or modelled curve. The marginal discrimination likely results from the fact that the dotted green line peaks, but then flattens off, meaning that at the highest level of ability, the discrimination is poorer. It is unclear from the item content why this would be the case.

[5] The number of items presented here differs slightly from that in Table 6.1 because some items were collapsed, and two items were excluded due to technical issues during administration.

Table 7.1 Item analysis for critical thinking

Item	Task	Aspect	Item estimate*	Item-total corr.	Infit
1	1	Aspect 2.2: Identifies assumptions and motivations	-0.882	0.40	1.05
2	1	Aspect 2.2: Identifies assumptions and motivations	0.010	0.47	0.99
3	1	Aspect 2.2: Identifies assumptions and motivations	0.853	0.39	1.05
4	1	Aspect 2.2: Identifies assumptions and motivations	0.523	0.54	0.91
5	1	Aspect 2.2: Identifies assumptions and motivations	1.057	0.50	0.93
6	1	Aspect 2.2: Identifies assumptions and motivations	1.427	0.43	0.97
7	1	Aspect 2.2: Identifies assumptions and motivations	0.554	0.53	0.93
8	2	Aspect 2.2: Identifies assumptions and motivations	0.709	0.34	1.07
9	2	Aspect 2.1: Applies logic	-1.574	0.43	0.99
10	2	Aspect 2.2: Identifies assumptions and motivations	-0.594	0.48	0.99
11	2	Aspect 2.1: Applies logic	0.407	0.44	1.16
12	2	Aspect 2.2: Identifies assumptions and motivations	-1.132	0.39	1.03
13	2	Aspect 2.1: Applies logic	-0.522	0.38	1.17
14	2	Aspect 2.2: Identifies assumptions and motivations	0.967	0.33	1.09
15	4b	Aspect 3.2: Evaluates options	-0.500	0.64	1.00
16	4b	Aspect 3.2: Evaluates options	-0.447	0.61	1.06
17	4c	Aspect 3.2: Evaluates options	-0.035	0.37	1.11
18	8a	Aspect 1.2: Discriminates amongst information	-0.854	0.42	1.02
19	8a	Aspect 1.2: Discriminates amongst information	-0.961	0.43	1.02
20	8a	Aspect 1.2: Discriminates amongst information	-0.657	0.48	0.97
21	8a	Aspect 1.3: Identifies patterns and makes connections	0.742	0.57	0.87
22	8a	Aspect 1.1: Identifies gaps in knowledge	0.783	0.52	0.93
23	9	Aspect 3.2: Evaluates options	0.051	0.63	1.04
24	9	Aspect 3.2: Evaluates options	0.065	0.68	0.95

* item-centred

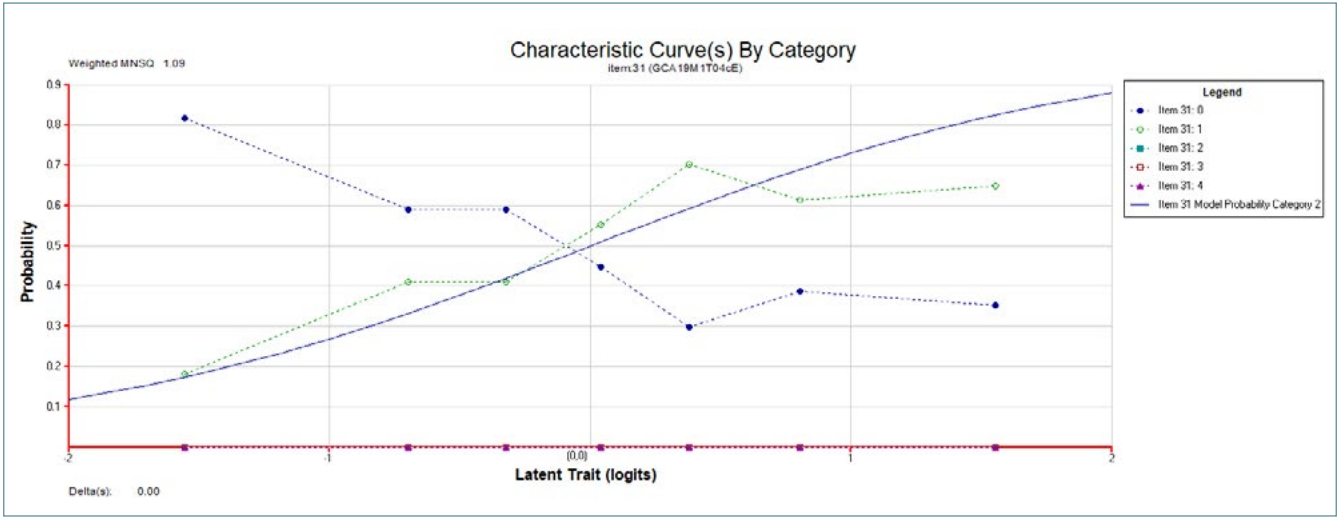


Figure 7.1 Item Characteristic Curve for Item 17 in critical thinking

Creative thinking

All nine items for creative thinking were calibrated together. A summary of the output is presented in Table 7.2. The item separation reliability is high at 0.998. This index indicates that items on the continuum were well separated and differentiated well between students' creative thinking abilities.

Table 7.2 Item analysis for creative thinking

Item	Task	Aspect	Item estimate*	Item-total corr.	Infit
1	3	Aspect 1.1: Number of ideas	-1.769	0.72	0.96
2	3	Aspect 1.2: Range of ideas	-0.884	0.71	0.90
3	3	Aspect 2.1: Shifting perspective	2.349	0.41	0.96
4	3	Aspect 3.2: Novelty	4.064	0.43	1.01
5	3	Aspect 3.1: Fitness for purpose	-2.370	0.58	0.95
6	3	Aspect 3.3: Elaboration	-0.375	0.56	0.97
7	10	Aspect 3.2: Novelty	0.201	0.46	1.04
8	10	Aspect 3.1: Fitness for purpose	-0.868	0.55	1.19
9	10	Aspect 3.3: Elaboration	-0.352	0.61	1.08

* item-centred

Almost all the items present a fit value within the confidence interval range with the exception of Item 8. This suggests that most items represented a single construct, interpreted as creative thinking, and provided good item quality in relation to measuring that skill.

In regards to Item 8, it was a partial credit item in Task 10, (the group assessment task), and purported to measure the extent to which the group's final refugee resettlement idea was fit for purpose. For a code 2, the group's idea needed to have been deemed both 'practical AND likely to be effective' by the markers (with code 1 = either/or). In comparison, Item 4 in Task 3 was also a measure of fitness for purpose of ideas, but in this instance the ideas were created individually, and had a good fit value (0.95). It is possible that the fitness for purpose manifests differently individually compared to a group, where identifying fitness of purposes for a group idea fits better with thinking critically about the idea.

Another explanation is that Item 4 measured fitness of purpose during the creative moment but Item 10 reflected on fitness of purpose for ideas that were generated several tasks before. It is possible fitness of purpose in relation to ccreative thinking can be measured well at the moment of idea creation, but as the fitness for purpose becomes more reflective this action moves towards being a critical thinking skill.

Collaboration

All 15 items for collaboration were calibrated together. A summary of the output is presented in Table 7.3.[6] The item separation reliability is high at 0.985. This index indicates that items on the continuum were well separated and differentiated well between students' collaborative abilities.

Table 7.3 Item analysis for collaboration

Item	Task	Aspect	Item estimate*	Item-total corr.	Infit
1	4c	Aspect 1.1: Communicates with others	-1.00	0.74	0.89
2	4c	Aspect 2.2: Recognises contributions of others	0.12	0.59	0.97
3	4c	Aspect 3.1: Ensures constructiveness of own contributions	-0.52	0.69	0.86
4	4c	Aspect 3.2: Resolves differences	-0.56	0.56	0.99
5	6	Aspect 1.1: Communicates with others	-1.00	0.34	1.21
6	6	Aspect 1.2: Pools resources and information	0.65	0.52	1.06
7	6	Aspect 1.3: Negotiates roles and responsibilities	2.58	0.46	1.13
8	6	Aspect 3.1: Ensures constructiveness of own contributions	-0.33	0.54	1.02
9	6	Aspect 3.2: Resolves differences	-0.04	0.58	0.96
10	8b	Aspect 1.1: Communicates with others	-1.77	0.48	1.11
11	8b	Aspect 2.2: Recognises contributions of others	0.62	0.57	0.97
12	8b	Aspect 3.1: Ensures constructiveness of own contributions	-0.32	0.58	0.92
13	8b	Aspect 3.2: Resolves differences	-0.12	0.49	1.01
14	8b	Aspect 3.4: Adapts behaviour and contributions for others	1.30	0.57	0.93
15	9	Aspect 3.3: Maintains shared understanding	0.38	0.60	1.00

* item-centred

[6] The number of items differs slightly from that in Table 6.1 because one item was not included in the IRT analysis.

Almost all the items present a fit value within the confidence interval range with the exception of Item 5. This suggests that most items represent a single construct, interpreted as collaboration, and provide good item quality in relation to measuring that skill. Item 5 measures students' abilities to communicate with others as indicated by at least one chat event either directing/clarifying the conversation (code 2) or responding to others (code 1). The explanation for the relatively poorer fit could relate to the fact that this item was recoded because no students presented with a code 0, so all students were communicating to some extent. It is possible that this is an example where communicating was necessary, but there was not sufficient exchange observed for it to be classified as an aspect of collaboration. Instead, perhaps what was observed was simply communication, and this specific example doesn't relate to their ability to collaborate.

Item 5 also presents low discrimination, which is shown in Figure 7.2 as a discrimination index between students who achieved a 1 and those who did not. This suggests that the item did not measure collaboration in the same way as the other items. A low correlation may also indicate a lack of variation in that the item was particularly easy or difficult for most students. This item likely shows low discrimination for the same reason that it shows poorer fit – that although one category defined in the scoring guide could be defended conceptually, it was not seen in this population of students and meant that the item had to be recoded.

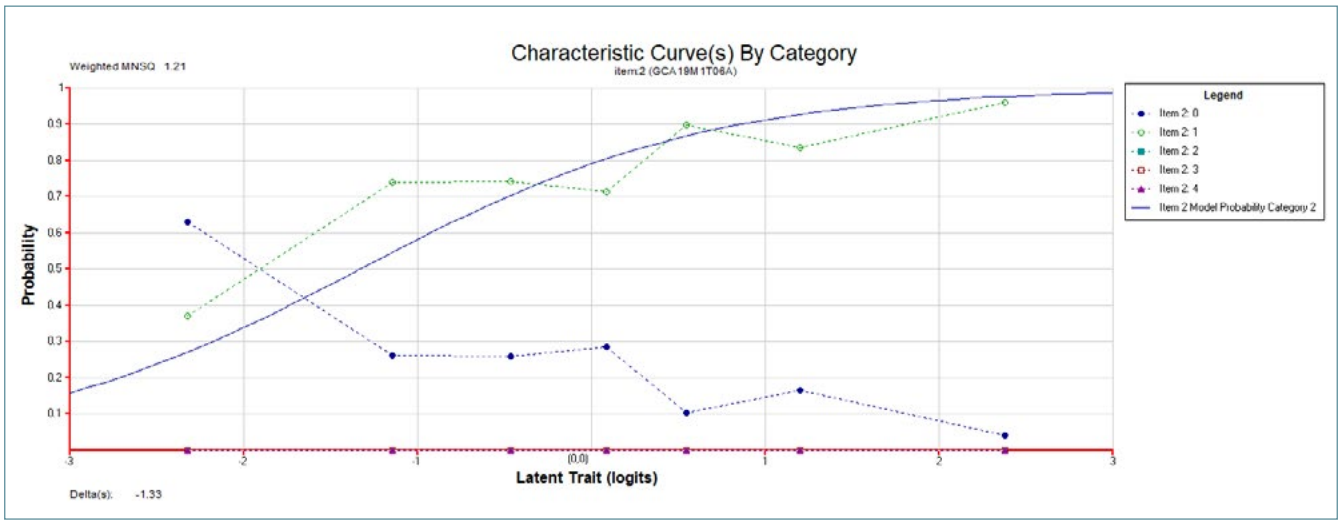


Figure 7.2 Item Characteristic Curve for Item 5 in collaboration

7.2 Distribution

An item's difficulty is partly identified through the frequency of the highest scoring response. If an item is difficult, then relatively fewer students should receive the highest score. Student success on a task is determined by the distance between the difficulty of the item and the ability of the student. For example, if the student is very capable and the task is easy, the probability of success is high; if the item is more difficult than the person is able, the probability of success is low.

It is important to check whether the items cover the range of item difficulty and student ability. This can be achieved by examining the item-person map, an output of the IRT analysis. This map places the item and student estimates onto a single scale, using logits as the scaling unit (an arbitrary unit used to enable location of the two variables on the same metric). It presents the items in increasing order of sophistication in relation to student ability and can be viewed as two vertical histograms displaying the spread of items. In order to interpret the scale, the item parameters are constrained with the mean item difficulty set at zero.

Critical thinking

The item-person map for the critical thinking items is presented in Figure 7.3 and indicates both the item difficulty and student ability were well distributed and aligned. The left side of the figure displays the distribution of student ability as a histogram ranging from roughly -3 to 2.5. The right side of the figure displays a slightly narrower distribution of the item difficulties ranging from about -1.8 to 1.4. Items 13.2 and 20 are at the top of the map, indicating they were the most difficult. Items 12 and 14.1 are at the bottom of the map indicating they were the easiest items. The student ability distribution extends lower than these items so it is difficult to discriminate between those students at this very low level, although there was only a small number of them.

Creative thinking

The item-person map for the creative thinking items is presented in Figure 7.4 and indicates both the item difficulty and student ability were well distributed and aligned. The left side of the figure displays the distribution of student ability as a histogram ranging from roughly -3 to 2.5. The right side of the figure displays a wider distribution, with the item difficulties ranging from about -4 to 4.5. Items 2.3, 3 and 6.2 are at the top of the map indicating they were the most difficult. Item 6.2 extends beyond the distribution of the student ability estimates, which suggests the item was too difficult for students. Items 1.1 and 2.1 are at the bottom of the map and were the easiest items. These items extend beyond the distribution of the student ability estimates suggesting these items were too easy for students.

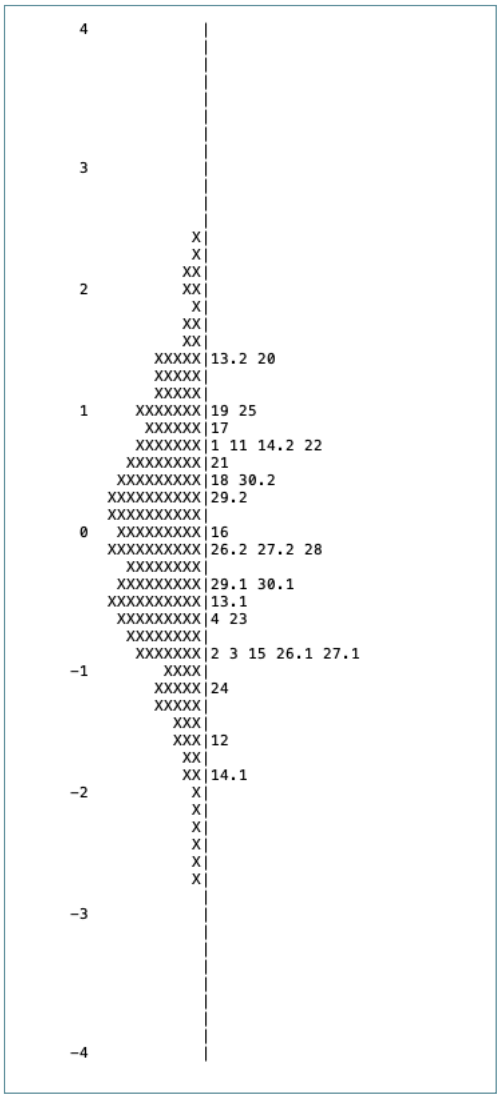


Figure 7.3 Item-person map for critical thinking

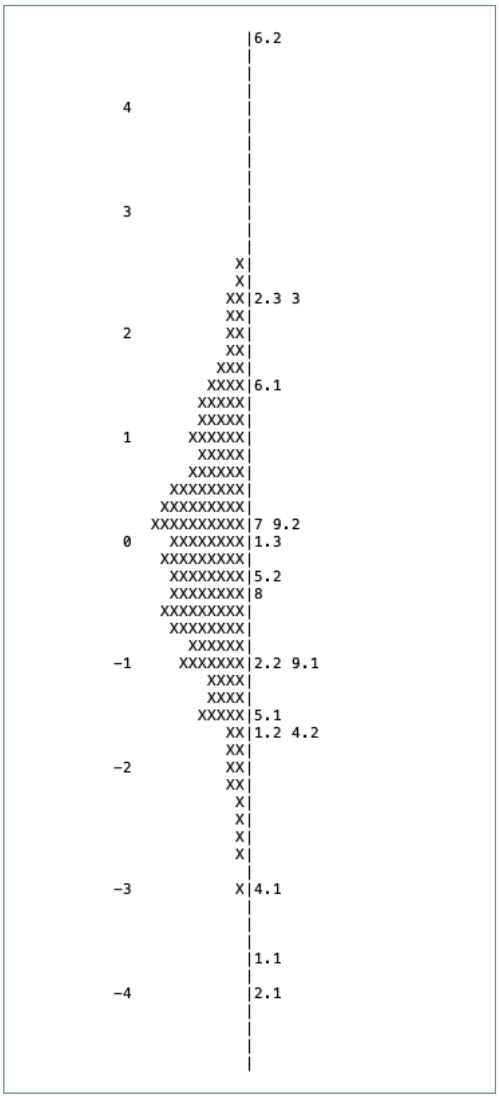


Figure 7.4 Item-person map for creative thinking

Collaboration

The item-person map for the collaboration items is presented in Figure 7.5 and indicates both the item difficulty and student ability were well distributed and aligned. The left side of the figure displays the distribution of student ability as a histogram ranging from roughly -4 to 3.5. The right side of the figure displays a slightly narrower distribution of the item difficulties ranging from about -3.2 to 2.6. Items 5.1 and 5.2 are at the top of the map indicating they were the most difficult. Item 3.1 is at the bottom of the map and was the easiest item.

7.3 Grade 5 versus Grade 8 comparison

A comparison between Grade 5 and Grade 8 students was completed. This analysis indicates whether there were differences in item performance by grade – that is, whether a Grade 5 student of a given ability performed the same way on the items as a

student of the same ability from Grade 8. Figures 7.6–7.8 show the scatterplot of item difficulties by grade for critical thinking, creative thinking and collaboration. In these scatterplots, items that remain in between the two blue lines show no real difference the way students of the same ability performed on the items between the grade levels. Items to the left of the two blue lines were easier for Grade 5s, while items on the right were easier for Grade 8s.

Two features of this analysis are noteworthy. First, the majority of items across all three domains showed no difference in the way students of the same ability performed depending on their grade level. Second, some items were, once adjusted for student ability, easier for Grade 8s (five items across the three skills). Other items were, once adjusted for ability, easier for Grade 5s (seven items across the three skills). The explanation for this is not clear from the item content. It is likely that this is simply an artefact of the relatively low number of Grade 5 students who completed the assessment.

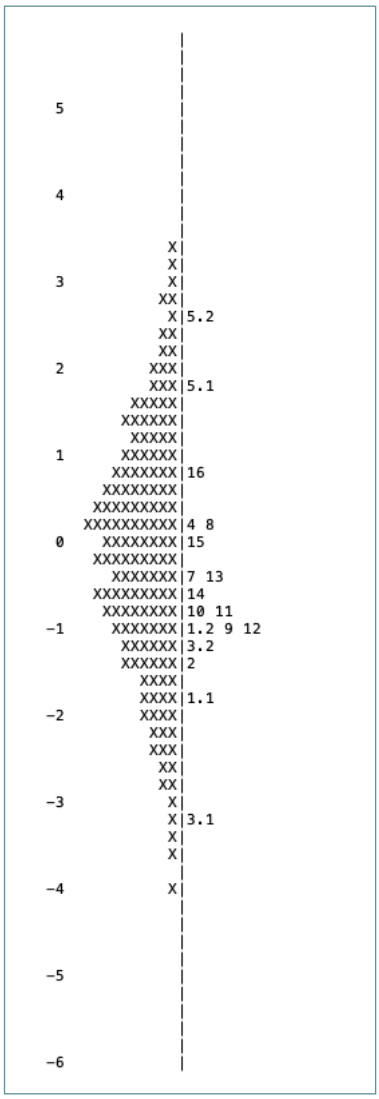


Figure 7.5 Item-person map for collaboration

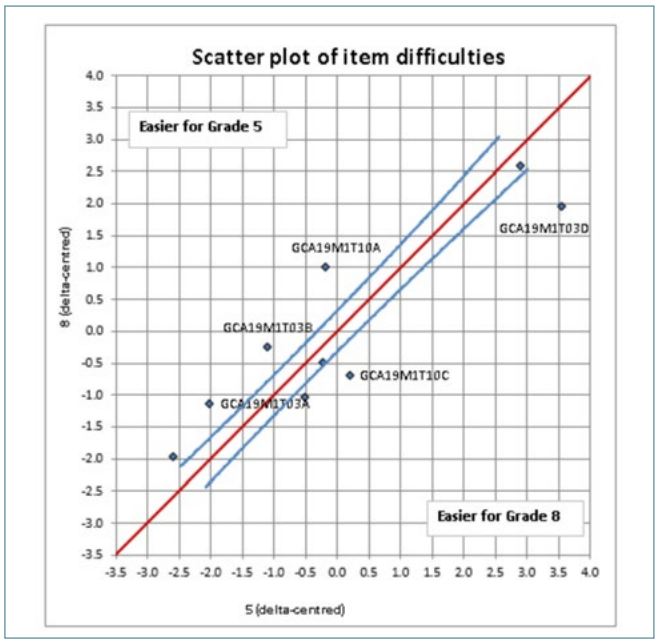


Figure 7.7 Scatterplot of item difficulties by grade for creative thinking

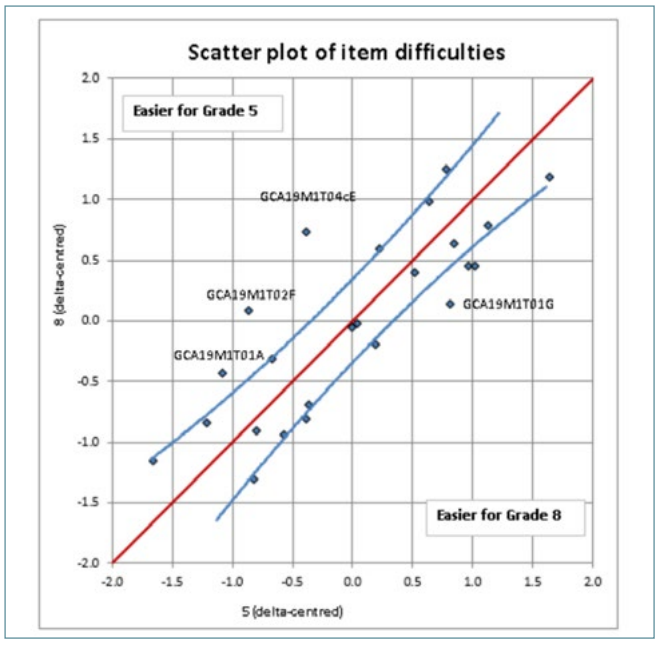


Figure 7.6 Scatterplot of item difficulties by grade for critical thinking

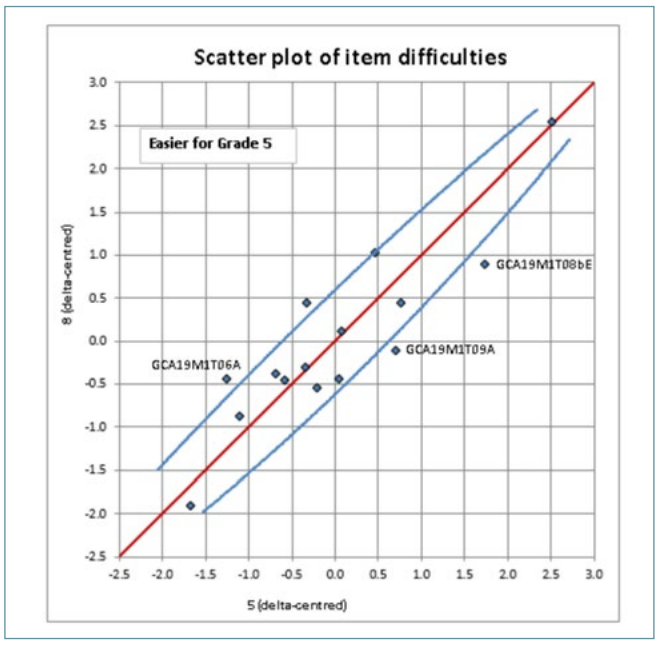


Figure 7.8 Scatterplot of item difficulties by grade for collaboration

7.4 Exploratory factor analysis

An exploratory factor analysis was performed to investigate whether certain groups of items fitted together.[7] Figure 7.9 shows the resulting scree plot. The number of factors are shown on the x-axis. This plot suggests that the best fit to the data is either a three- or a four-factor structure since it is after this point (after the line has passed 3 or 4 on the x-axis) that the line flattens off, suggesting that a little more explanation of the way the items grouped together is offered by adding more factors. In the context of this assessment, which was intended to measure three skills, a three-factor structure makes intuitive sense. Indeed, a close examination of which items grouped together in both the three- and four-factor models showed that the assessment measured three skills.

In the three-factor model, Factor 1 might be labelled 'collaboration', with all but two items from this skill loading on this factor. Factor 2 might be labelled 'critical thinking', since more than two-thirds of the critical thinking items load onto this factor (alongside some creative thinking items). Factor 3 might be labelled 'creative thinking' – again, two-thirds of the creative thinking items load onto this factor alongside a small number of critical thinking items. This modelling does seem to suggest that critical and creative thinking are related to each other, which is consistent with the way these skills have been described by some practitioners (e.g. ACARA, n.d.). A close item analysis offers further explanation in some cases. For example, two of the creative thinking items that loaded on the factor identified as critical thinking were about elaborating an idea. While the research literature clearly identifies this as part of creative thinking, it is not difficult to see how this skill might draw on some skills related to critical thinking.

The four-factor structure also offers support for the notion that critical and creative thinking, while separate skills, have a strong relationship with each other. In this four-factor model, Factor 1 was clearly collaboration, since it contained all but two of the

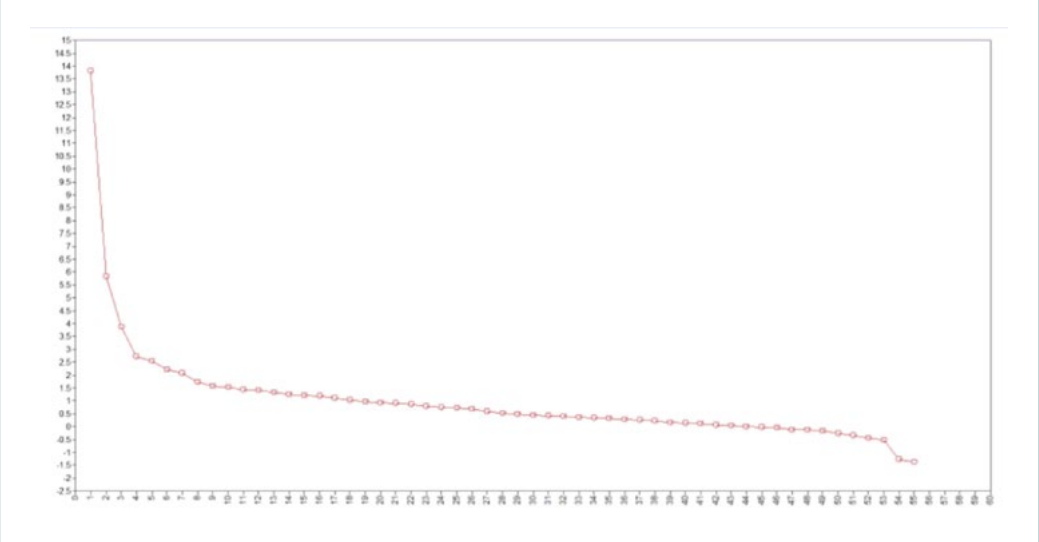


Figure 7.9 Exploratory factor analysis scree plot

[7] Because of technical issues during the assessment, this factor analysis shows that responses that were presented but not answered were treated as not administered (rather than missing).

collaboration items. Factor 2 comprised a subset of about one-third of the critical thinking items, and no other items. Every item in this factor drew on a common skill: all items in this factor came from Task 8a, and focused on conducting research and using the results to improve an idea. Factor 2, then, represents a subset of critical thinking. Specifically, this factor contained 9 of the 11 items that focused on Strand 1 – knowledge construction – and so supports not only the notion that the assessment taps into critical thinking, but also for the specific framework structure proposed. Factor 3 contained two-thirds of the creative thinking items, as well as almost all remaining critical thinking items (16 items) that were not present in either Factor 1 or 2, again suggesting that there was a relationship between these two skills. Finally Factor 4 contained the remaining three creative thinking items that were not included in any other factor.

7.5 Developing understanding about the skills

What does the assessment tell us about the skills? Assessment data should provide sufficient information about a construct so that understanding can be improved and judgements can be validated. The assessment data provided validation and iteratively informed the definitions in the skill development framework and the levels of skill development. The following task examples and item examples highlight this.

Three tasks from the humanities assessment module, one for each skill, are presented. This demonstrates the assessment content and context, as well as the type of data being elicited from the tasks in relation to each of the skills.

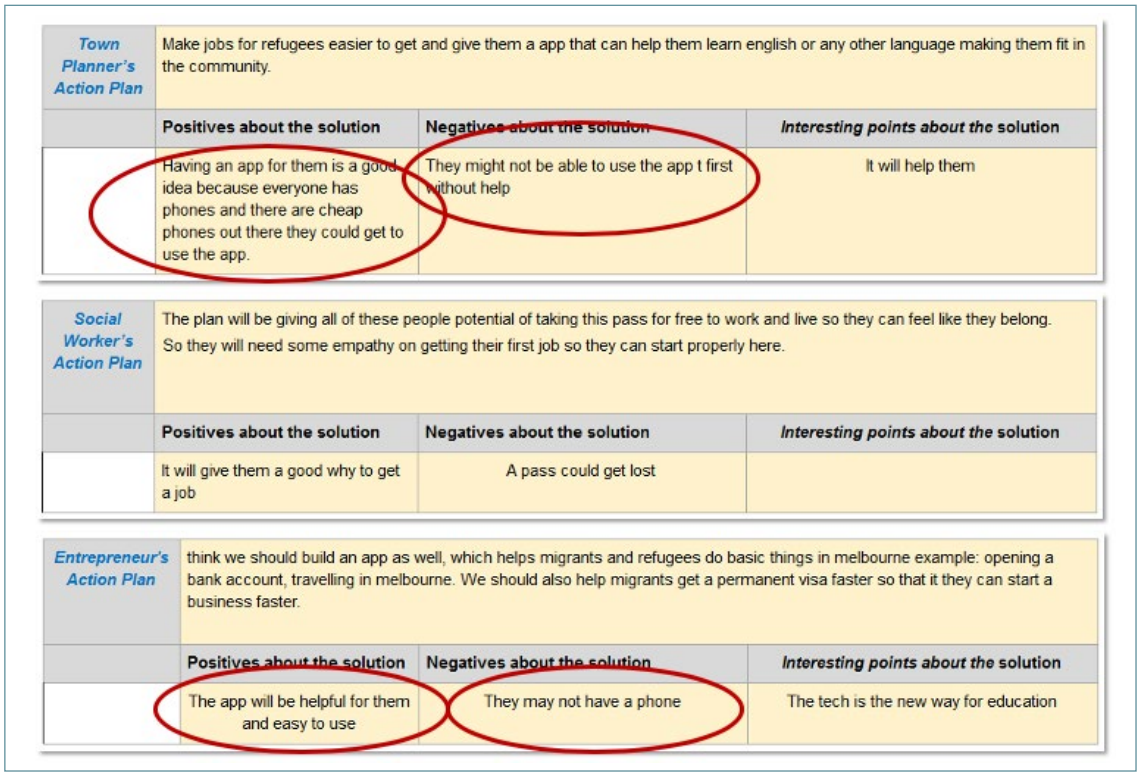
Critical thinking: Task 8 – Evaluation of group's action plan

Task 8 measured critical thinking through an evaluation and critique of the groups' community action plan. After deciding on roles, the task directed group members to read/view/listen to a different set of resources on refugee resettlement. They made research notes from these resources, before devising a *new* possible solution to the problem of how to smoothly resettle refugees in the local area. They were instructed to consider the information they just read in the formulation of their new idea, to elaborate on their idea and to pose further questions they think they would need to know answers to, to ensure the success of their idea. In Task 8, the students were directed to critically evaluate their own and each other's solutions. Table 7.4 presents the mapping of Task 8 to the critical thinking skill development framework to give an insight into how the task was scored.

Table 7.4 Mapping of Task 8 to the critical thinking skill development framework

		Mapping to the skill development framework		
Item	**Description**	**Skill**	**Strand**	**Aspect**
8A	Finds relevant information in a resource	Critical thinking	1. Knowledge construction	1.2. Discriminates amongst information
8B	Explains how identified information supports an idea	Critical thinking	1. Knowledge construction	1.3. Identifies patterns and makes connections
8C	Identifies gaps in knowledge	Critical thinking	1. Knowledge construction	1.1. Identifies gaps in knowledge

Figure 7.10 presents a high scoring student response to Task 8. This student scored high on the criteria in Table 7.4, which indicates higher proficiency in critical thinking. One clear positive response was offered about another team member's ideas, but it did not clearly explain a specific positive outcome of the idea (the ubiquity of phones was not sufficiently explained as a positive outcome of the solution; the 'cheapness' of phones was fairly relative). One negative was offered about another team member's ideas that related to a possible and specific real-world constraint of the solution (that the app in and of itself might not be self-explanatory to use). The response 'a pass could get lost' was not a constraint of the solution per se; assuming one doesn't lose their pass, what was a constraint of the solution of having a 'pass' itself. The positive identified in their own solution was a simple, circular summary of elements already stated in the solution (a positive of an app 'which helps migrants' is that 'the app will be helpful'). The negative identified in their own solution refered to specific limitations arising from the solution itself – it could not be a solution if migrants don't have/can't afford a phone.



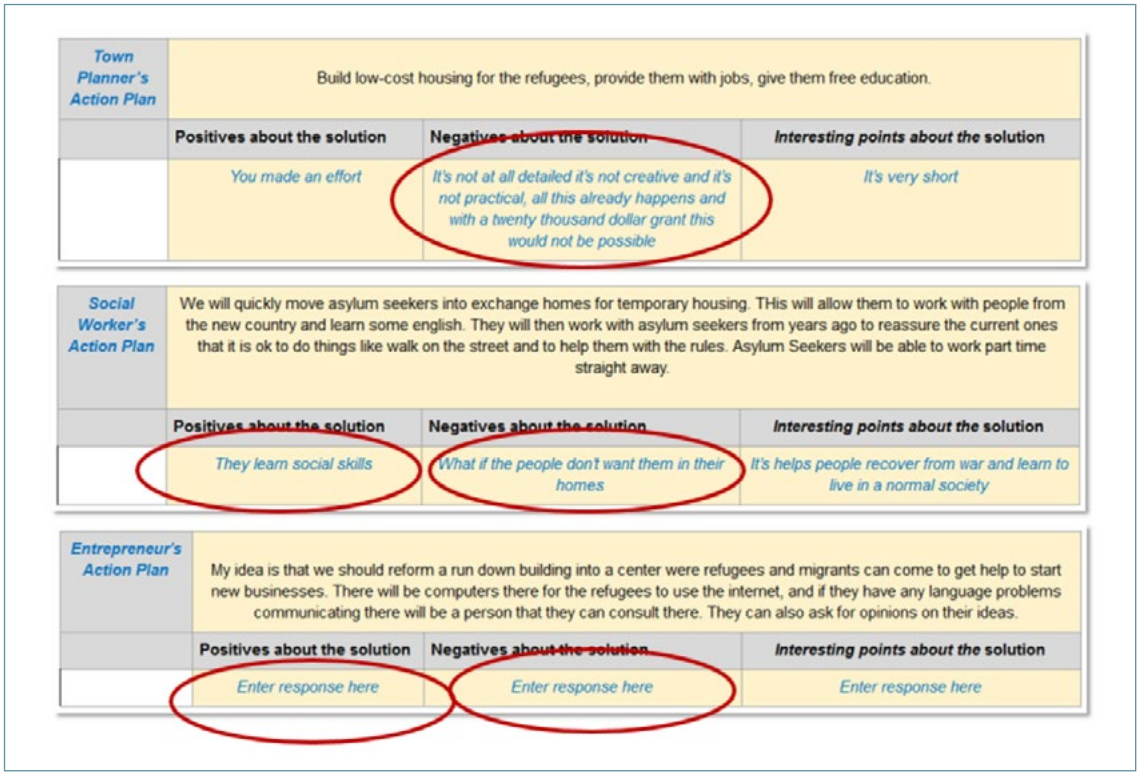
Town Planner's Action Plan	Make jobs for refugees easier to get and give them a app that can help them learn english or any other language making them fit in the community.		
	Positives about the solution	**Negatives about the solution**	***Interesting points about the* solution**
	Having an app for them is a good idea because everyone has phones and there are cheap phones out there they could get to use the app.	They might not be able to use the app t first without help	It will help them

Social Worker's Action Plan	The plan will be giving all of these people potential of taking this pass for free to work and live so they can feel like they belong. So they will need some empathy on getting their first job so they can start properly here.		
	Positives about the solution	**Negatives about the solution**	***Interesting points about the* solution**
	It will give them a good why to get a job	A pass could get lost	

Entrepreneur's Action Plan	think we should build an app as well, which helps migrants and refugees do basic things in melbourne example: opening a bank account, travelling in melbourne. We should also help migrants get a permanent visa faster so that it they can start a business faster.		
	Positives about the solution	**Negatives about the solution**	***Interesting points about the* solution**
	The app will be helpful for them and easy to use	They may not have a phone	The tech is the new way for education

Figure 7.10 Example of high scoring student response to Task 8

Figure 7.11 presents a low scoring student response to Task 8. This student scored low on the criteria in Table 7.4, which indicates lower proficiency in critical thinking. Only one positive response was offered about another team member's ideas that related to an outcome of the solution. Two negatives were offered about another team member's ideas that related to possible constraints of the solution. No positives or negatives were identified in their own solution. The critique was arguably much more thoughtful and coherent than the higher-scoring student's. Had the lower scoring student completed an evaluation of their own solution they would have scored higher.



Town Planner's Action Plan	Build low-cost housing for the refugees, provide them with jobs, give them free education.		
	Positives about the solution	Negatives about the solution	Interesting points about the solution
	You made an effort	It's not at all detailed it's not creative and it's not practical, all this already happens and with a twenty thousand dollar grant this would not be possible	It's very short
Social Worker's Action Plan	We will quickly move asylum seekers into exchange homes for temporary housing. THis will allow them to work with people from the new country and learn some english. They will then work with asylum seekers from years ago to reassure the current ones that it is ok to do things like walk on the street and to help them with the rules. Asylum Seekers will be able to work part time straight away.		
	Positives about the solution	Negatives about the solution	Interesting points about the solution
	They learn social skills	What if the people don't want them in their homes	It's helps people recover from war and learn to live in a normal society
Entrepreneur's Action Plan	My idea is that we should reform a run down building into a center were refugees and migrants can come to get help to start new businesses. There will be computers there for the refugees to use the internet, and if they have any language problems communicating there will be a person that they can consult there. They can also ask for opinions on their ideas.		
	Positives about the solution	Negatives about the solution	Interesting points about the solution
	Enter response here	Enter response here	Enter response here

Figure 7.11 Example of low scoring student response to Task 8

Creative thinking: Task 3 – Generating ideas

Task 3 measured creative thinking and requires students to generate creative ideas to solve a problem. Prior to this, the first two tasks involved analysing an FAQ centred on the main problem-solving scenario and a critical analysis of opinion pieces in relation to the problem-solving scenario. Table 7.5 presents the mapping of Task 3 to the creative thinking skill development framework to give an insight into how the task was scored.

Figure 7.12 presents an example of a student response to Task 3. This student scored low on the criteria in Table 7.5 indicating less proficiency in creative thinking. The student showed limited fluency in their ideas, and not all ideas were stated as a coherent solution. The ideas showed naivety about the problem context, which limited their effectiveness as solutions. There was a lack of detail and the solutions suggested limited perspectives were considered.

Table 7.5 Mapping of Task 3 to the creative thinking skill development framework

		Mapping to the skill development framework		
Item	**Description**	**Skill**	**Strand**	**Aspect**
3A	Develops number of ideas	Creative thinking	1.Generation of ideas	1.1. Number of ideas
3B	Develops a range of ideas	Creative thinking	1.Generation of ideas	1.2. Range of ideas
3C	Develops ideas that consider multiple perspectives	Creative thinking	2.Experimentation	2.1. Shifting perspective
3D	Develops ideas that are novel	Creative thinking	3.Quality of ideas	3.2. Novelty
3E	Develops ideas that are fit for purpose	Creative thinking	3.Quality of ideas	3.1. Fit for purpose
3F	Develops ideas that are elaborated	Creative thinking	3.Quality of ideas	3.3. Elaboration

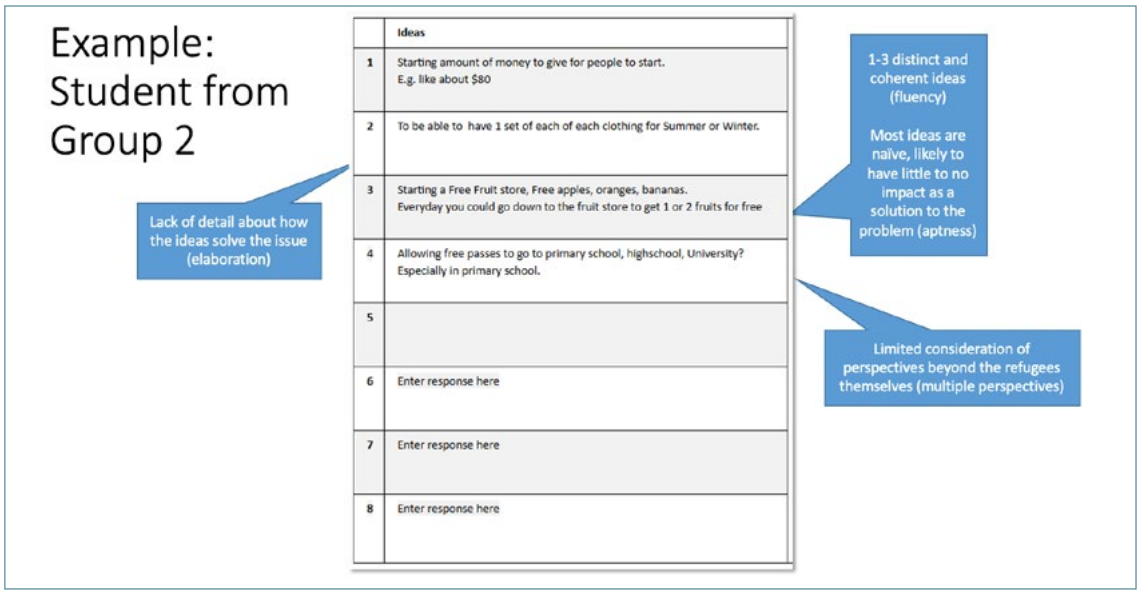


Figure 7.12 Example of low scoring student response to Task 3

Figure 7.13 presents a second example of a student response to Task 3. By contrast with Figure 7.12, this student scored high on the criteria in Table 7.5 indicating more proficiency in creative thinking. This student showed a fluency of ideas which were mostly apt, or fit for the purpose, of assisting refugees to resettle smoothly. The ideas were distinct from each other and showed a degree of flexible thinking. While quite diverse, they mostly hit on some expected solutions – making refugees feel welcome, providing housing solutions and education services, personal assistance, etc. However, a novel idea emerged about creating host families in the local community. Ideas that discussed bringing the wider community together rather than simply providing goods and services to the refugees also showed the student considered multiple perspectives. The student could elaborate their reasoning about the outcome of an idea and therefore scored high on this specific criterion.

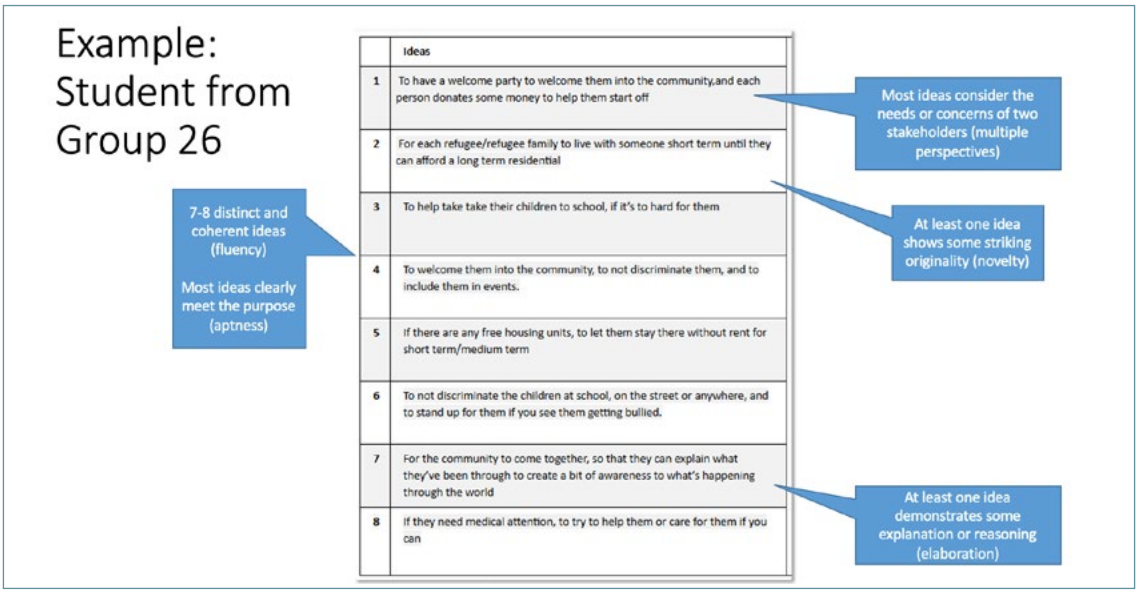


Figure 7.13 Example of high scoring student response to Task 3

Collaboration: Task 6 – Role agreement

In Task 6, students used a group chat platform to negotiate their role in the activity. Each member of the group selected a different role, which then determined the information they had access to. Each student was required to review each of the role descriptions, decide which role they wanted to negotiate for themselves, and the reasons why (see Figure 7.14). Students were also encouraged to think about the other roles and why they wouldn't be the best person suited to that role.

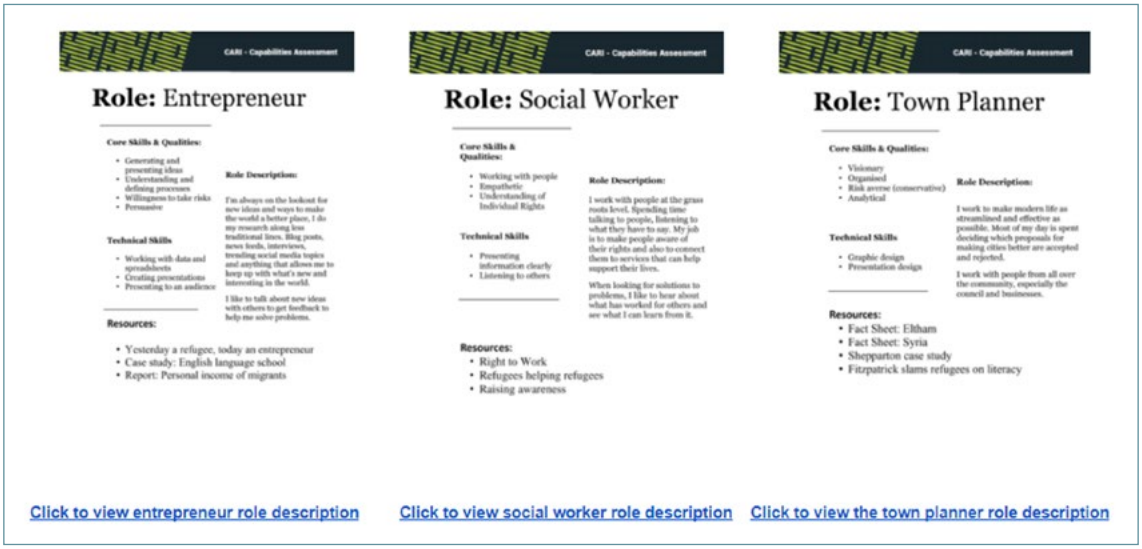


Figure 7.14 Role options presented to students

After the roles were assigned, students had 10 minutes to negotiate. They entered a Google Hangouts chat space to start a group chat with their group members. At the end, they documented which team member was allocated to each role, and why. Table 7.6 presents the mapping of Task 6 to the collaboration skill development framework to give an insight into how the task was scored.

Table 7.6 Mapping of Task 6 to the collaboration skill development framework

		Mapping to the skill development framework		
Item	**Description**	**Skill**	**Strand**	**Aspect**
6A	Communicates with group for the allocation of roles	Collaboration	1.Building a shared understanding	1.1. Communicates with others
6B	Shares information for decision-making in group	Collaboration	1.Building a shared understanding	1.2. Pools resources and information
6C	Regulates contributions to group to effectively allocate roles	Collaboration	3.Regulating	3.1. Ensures constructiveness of own contribution
6D	Resolves differences in group to allocate roles	Collaboration	3.Regulating	3.2. Resolving differences
6E	Negotiates roles with group members	Collaboration	1.Building a shared understanding	1.3. Negotiates roles and responsibilities
6F	Adapts contributions to group members for the allocation of roles	Collaboration	3.Collectively contributing	3.4. Adapts behaviour and contributions for others

Figure 7.15 presents a sample extract from Group 2's conversation, which is an example of a low scoring group response to Task 6. They were staying on-task, and so showed perseverance and regulated their social interaction. They responded to each other's contributions, they initiated threads of conversations, some justified their ideas, and yet there was an overall disjointedness and a lack of depth of development or conflict resolution to the conversation.

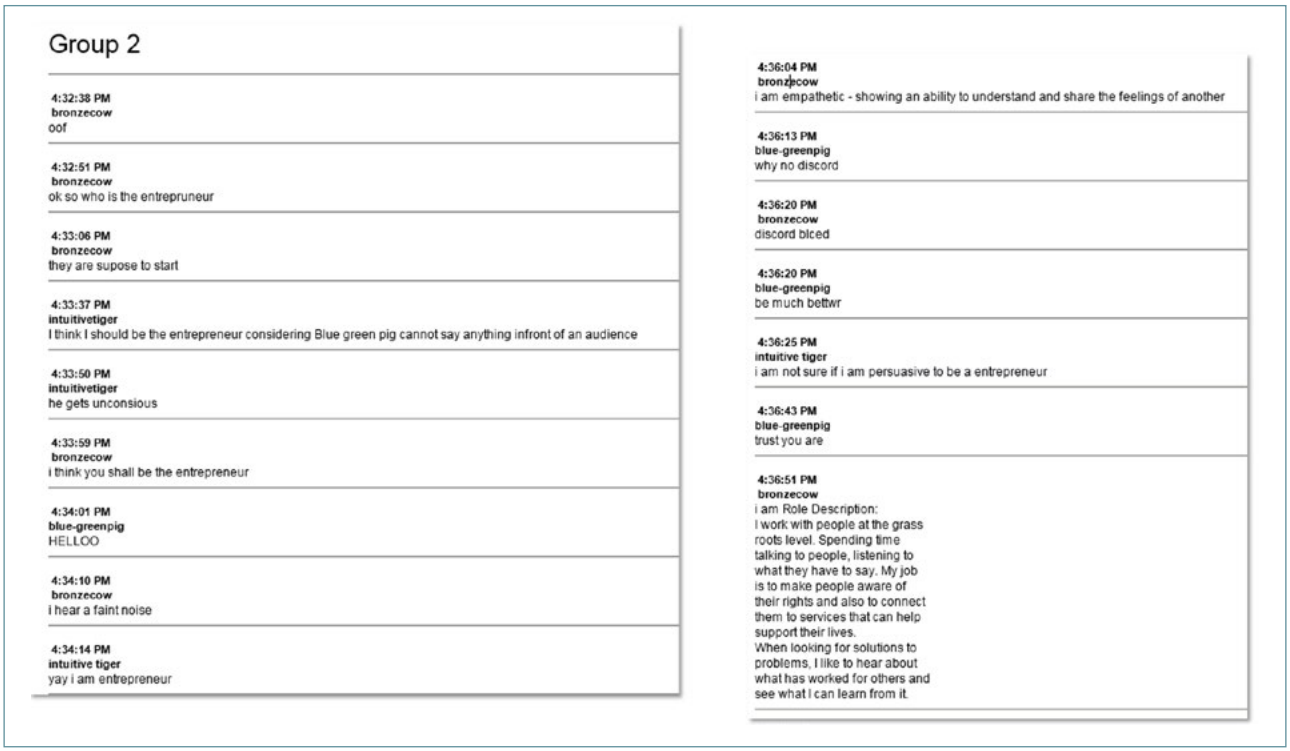


Figure 7.15 Example of low scoring group response to Task 6

In comparison, Figure 7.16 presents a sample extract from Group 26's conversation, which is an example of a high scoring group response to Task 6. This conversation does not reveal much conflict resolution, and arguably less discussion of their thinking about which role to take than Group 2, but there was a coherence to the conversation based around a problem to be solved, about what to do in the activity and how to do it. 'Modest Alligator' was least involved, but the other two team members were very responsive in ensuring the team knew what they were doing and achieved the right outcome despite confusion.

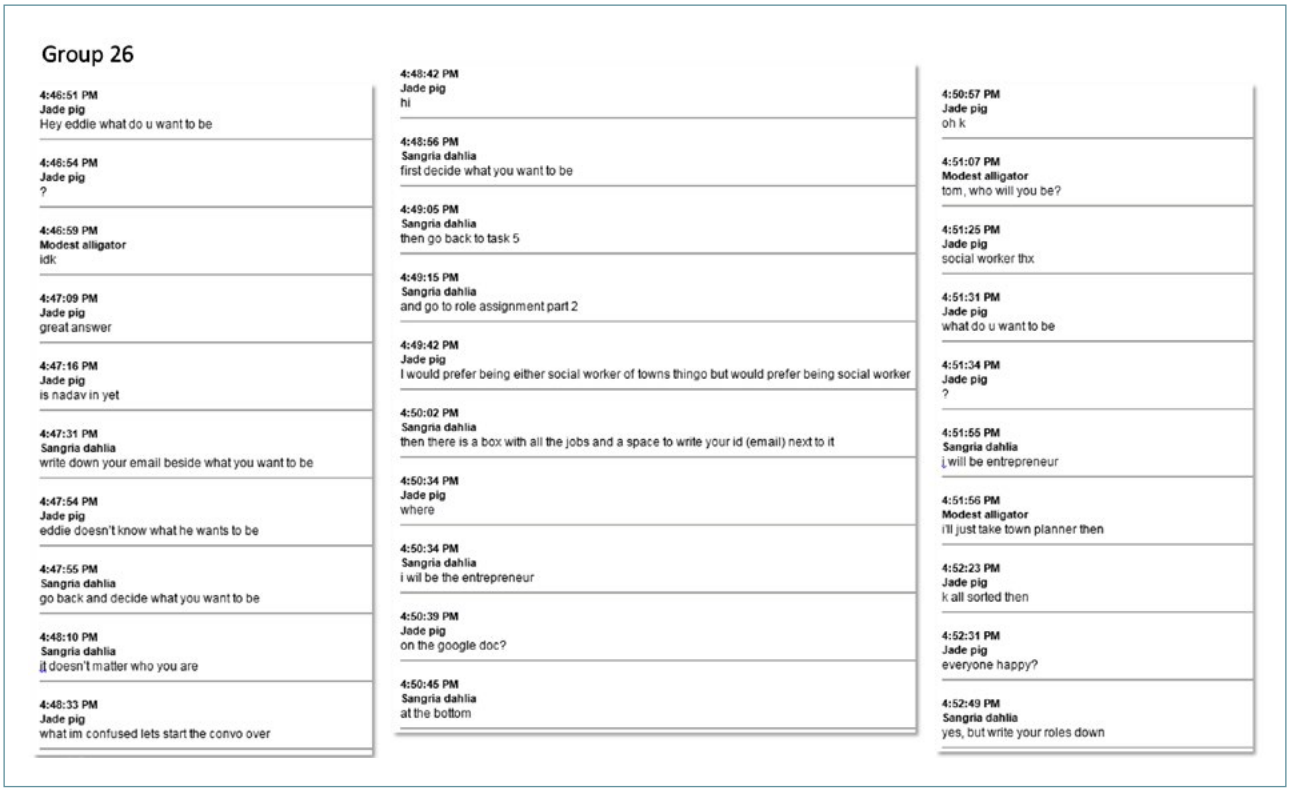


Figure 7.16 Example of high scoring group response to Task 6

7.6 Item examples

Critical thinking

Despite the relatively high number of critical thinking items within the assessment, it is important to acknowledge that this did not achieve complete coverage of the construct. Three aspects of critical thinking (2.3: Justifies arguments; 3.1: Identifies criteria for decision-making; and 3.3: Tests and monitors implementation) were unassessed. Further to this, the nature and the constraints of the assessment meant that the six aspects of critical thinking that were assessed were not evenly covered; indeed, over 80 per cent of critical thinking items assessed just three aspects (1.2: Discriminates information; 2.2: Identifies assumptions and motivations; and 3.2 Evaluates options). Any consideration of what this assessment reveals about critical thinking, therefore, needs to be couched within these caveats.

That said, the item difficulty map shown in Figure 7.3 tentatively suggests some interesting differences between the aspects of critical thinking. The vast majority of the most difficult items were those that assessed Strand 2: Evaluates Reasoning (from Task 1 and Task 2). Specifically, those items that required students to apply logic to identify a conclusion that would follow from a stated opinion, or to identify likely motivations or assumptions 'behind' an opinion or frequently asked questions. This may indicate that inference – a key component of reasoning – is more difficult than some other facets of critical thinking such as the identification of information relevant to a given context (such as was required in Task 8a), or explaining the pros and cons of a solution to a given problem (such as was required in Task 4b). In both such tasks inference is certainly still applicable, but perhaps less abstractly so.

Some multiple-choice items in Task 2, such as Items 13 and 14 potentially shed some interesting insight into the development of logical reasoning. These items ask candidates to identify a conclusion that would follow from a stated opinion, and upon initial analysis both appeared to have competing keys (the average ability of students choosing the two options was similar). In both cases, the apparent competing key was not strictly speaking a conclusion that followed logically from an opinion. Though unlike the other available distractors, it was a proposition *not inconsistent with* the given opinion: a necessary (but not sufficient) condition for a conclusion. As a result, these items were subsequently re-scored using partial credit. This result may suggest that identifying propositions consistent with another is a crucial first developmental step towards learning how to identify a logical conclusion of a proposition.

Creative thinking

If we review the item difficulty map for creative thinking shown in Figure 7.4, it is of particular interest that unlike the other two skills, there were items that had a difficulty level outside the range of student ability present in the sample of students. That is, there were two items that were too easy for the sample of students, and one item that was too difficult. It is interesting to examine the features of the hardest and easiest items in the assessment as this has the potential to develop understanding about the skill.

The two easiest items and the most difficult item were in Task 3, which focused on developing a set of creative ideas in response to a problem. Initially, the set of ideas developed by the students was scored on how many distinct ideas they were able to generate, and whether the ideas represented a range. Each of these were scored using a partial credit model, where the lowest partial credit score for 'number of ideas' signified that the student generated a small number of ideas and the lowest partial credit score for 'range of ideas' suggested that the type of ideas were distinctly different from one another. It is these partial credit score points that were the easiest creative thinking items in the assessment. This suggests that even for students with low proficiency in creative thinking, it is not difficult to generate a small number of distinct ideas. More generally, this finding suggests that the stimulus material was at an appropriate level: if students were unable to engage with the problem situation, even generating a small number of ideas would have been beyond them. It is worth noting that in the course of administering the assessment, some teachers commented that they felt that the topic of refugee resettlement was too complex for Grade 5 students. The results suggest that it is possible for younger students to engage with

sophisticated social problems if they are given the appropriate scaffolding, and the problems are framed appropriately.

The most difficult creative thinking item in the assessment also related to the generation of a set of ideas to solve a problem, but this time, novelty was the criteria scored. Students who were able to give at least one idea that was striking in its originality received credit for this item. Novelty was considered relatively to the test population – it is unlikely that a student would generate a novel idea in the sense of that idea never having been generated before. Nevertheless, generating at least one novel idea was beyond the ability of the students in this sample. This makes intuitive sense: a key feature of creative thinking is the ability to generate original ideas, but truly original ideas are difficult to generate. This simple review of Task 3 provides information about the difficulty of different aspects of the skill, consideration of which will inform future research.

Collaboration

It is acknowledged that there were some opportunities for collaboration that were not captured by the assessment. However, three tasks were designed to facilitate collaboration and capture scorable data. All three of the collaborative tasks had the same structure and method for capturing data. These took the form of a group text chat tool where students were required to collectively make a decision. The first collaborative task required students to choose one idea from the ideas submitted by the three members of the group. The second task required students to negotiate a pre-defined role that determined what resources they could access in the next task and was designed so that the roles had certain responsibilities that would be perceived as favourable or unfavourable depending on the students' perspectives. The third task required students to bring independent research into a discussion to improve on and finalise their group's idea. For the most part, the assessment criteria were the same for each of the three tasks.

In Figure 7.5, Items 2, 1.1 and 3.1 each represent a basic form of communication from each of the three tasks. These items were the easiest for students and this suggests that most students can communicate in a collaborative environment irrespective of the purpose of the collaboration. The higher score categories for Items 3.2 and 1.2 represented more sophisticated communication to demonstrate the capacity to direct conversations as opposed to simply responding to the direction of others.

Items 9, 10 and 11 represent the ability to regulate one's own contributions in a collaborative environment. This was scored by comparing the number of relevant contributions with the number of irrelevant or distracting contributions. Regulating one's own contributions appears to be more difficult than communicating, and, as we would expect, these items were clustered together in terms of their difficulty despite being separate items from separate collaboration tasks.

Items 12, 13 and 14 represent the capacity to make attempts to resolve differences or conflicts (irrespective of whether they were in fact resolved) that naturally arose from the tasks as a result of trying to choose one idea over another, negotiating roles or trying to finalise the group's idea. These items are also clustered together with respect

to their difficulty; however, it appears that resolving differences when deciding roles (Item 13) was significantly more difficult than resolving differences when deciding on the best ideas (Item 12). This might suggest that students have more experience discussing ideas and have developed strategies and protocols for resolving differences when deciding on the best ideas but when it comes to role allocation there may be a stronger attachment to the outcome, or perceived implications, making it more difficult to make attempts at resolving differences.

At the upper end of the difficulty range, Items 5.1 and 5.2 also related to the role negotiation task. These items represent the ability of students to employ strategies for assigning roles such as obtaining information about a group member's skillset or personal interests and matching it with the description of a role and its associated responsibilities. This suggests that the most difficult aspect of collaboration in the assessment was the employment of diplomacy when members of the group are emotionally invested in an outcome contrary to your own.

8 MEASUREMENT CONSIDERATIONS

8.1 Quality control of marking processes

Seven of the tasks were scored by markers who had undergone an extensive training process. Many of the tasks contained multiple parts, where each part was called an item. Consequently, markers scored a total of 93 different items across the two modules.

Given the innovative nature of the material, numerous quality control measures (in addition to the extensive training) were implemented both during, and subsequent to the main marking period. These included spot checking, discrepancy identification and data adjudication.

Spot checking

During the course of the marking, an experienced lead marker conducted spot checking of other markers' work, updating scores and conducting supplementary training with individual markers as needed. Overall, ten per cent of scores were spot checked.

Discrepancy identification

In the course of the marking period, all student responses were double-marked – that is, were marked separately by two markers from the team. Therefore, once marking has been completed, it is possible to identify the levels of discrepancy that were present for each task.

Discrepancy rates ranged from 3–58%[1] for module 1, and 7–39% for module 2, as provided in Appendix B.

Data adjudication

Calculating the discrepancy rate for each item allowed the identification of items to be targeted for data adjudication. A strict approach to coder reliability was taken, with almost half of the items undergoing some re-marking. The items with the highest discrepancy rates were re-marked by a leading marker, who gave a definitive score that was then used for the final analyses. In all cases, at least half of the discrepant responses were re-marked, and in some cases all discrepant responses were re-marked.

Observations

Two observations are warranted.

First, *in general*, items seemed to have similar coder reliability across both modules that were based upon different learning area contexts. Task 4b, for example, was a clear candidate for data adjudication in both modules, while, conversely, the majority of items in Task 8a were able to be marked reliably across both modules. This is important information for test developers to review when considering revisions to the coding guides.

A second observation is that the items seen here are, particularly in the context of a new and innovative assessment, work very well. To give a sense of what is an acceptable level of coder reliability, consider the metric used in the large-scale international assessments Trends in International Mathematics and Science Study (TIMSS) and Progress in International Reading Literacy Study (PIRLS). In these influential assessments, agreement above 85 per cent (or a discrepancy rate of less than 15%) is considered good, while agreement of above 70 per cent (or a discrepancy rate of less than 30%) is considered acceptable. If we apply this metric to general capabilities assessment results, 38 of the 93 items had good agreement, and a further 39 had acceptable agreement. In total, 77 of the items would, under the TIMSS/ PIRLS rule, not have needed data adjudication. In practice, a more rigorous approach was taken for this project, with some of these 77 items undergoing some level of re-marking. Taken together, the quality control processes implemented before, during, and after the marking processes mean there can be a high degree of confidence in the results obtained.

8.2 Local independence

Local independence is one of the fundamental assumptions of the one-dimensional IRT model and consists of two parts: local item independence and local person independence (Reckase, 2009). That is, within an assessment, student responses to items should not, after ability is considered, have any relationship with each other (Embretson & Reise, 2000). In regards to managing local item independence, all of the items were designed to be independent of one another. Scoring criteria do not depend on criteria in other items, or on the presence or absence of another item. Given the nature of the tasks, particularly the collaborative nature of some, local person independence is more complex. In cases where the item was not relevant to the student role being scored, or in cases where the behaviours would be directly

dependent upon the other student, these would be the equivalent of the item not being administered, and was therefore recorded as a missing value. These scoring rules are placed in order to not violate the assumptions of local independence. However, the very nature of collaboration requires dependence, therefore additional psychometric models could be applied to investigate this.

8.3 Group scores

When assessing something that is collaborative, an immediate issue is whether an individual should be assessed within a group, the group assessed as a whole, or both. The general capabilities assessment contained some activities that assessed individuals, and others in which a group score was given. While this may seem an ideal solution, it presents its own challenges in relation to the choice of statistical model. IRT models, for example, assume local independence, that is, that a test-taker's response to an item is independent of the response of other test-takers. If group scores are given, by definition this assumption is violated. This observation reinforces the point that assessing skills may require new and more innovative approaches to both assessment and scoring.

9 FINDINGS AND CONCLUSION

The ACER general capabilities assessment has demonstrated that critical thinking, creative thinking and collaboration can be assessed using the approach adopted, that is, multiple skills assessed within domain-orientated problem-based tasks. The assessment data also contribute to validation of ACER's skill development frameworks and their accompanying levels of skill development.

The advancement of this field of research will only occur if embedding the skills can be aligned across curriculum, assessment and pedagogy. Further research will focus on strengthening these links and achieving this will require an iterative process of refinement and validation of embedding skills within and across each alignment component.

9.1 Revisiting alignment

As an underlying principle in this work, alignment needs to be ensured but one must begin somewhere. This project started with assessment first, but that has an impact for all components of alignment: curriculum, future assessment work, and pedagogy.

9.2 Implications for assessment

Refinement of assessment platform

The assessment tools in this project are currently prototypes trialled across Australia. Analysis of the trial data enables researchers to refine the delivery and resourcing, so as to ensure valid, accurate and well-targeted assessment of the general capabilities.

Further data collection and analysis can allow for deeper investigations, for example:

- How do individual students perform across the different skills when measured simultaneously in the same context?
- What is the dependency between the skills? How can this help us understand how to teach them?
- How do the skills develop from Grade 5 to Grade 8? What does the rate of growth look like for these skills?
- How, if at all, do the skills transfer across different learning areas?

While the approach taken in this project was beneficial in gaining understanding of the skills, the platform used was inefficient in relation to administration time and the limited access to process data. Additional formative assessment tasks should be developed that allow students to demonstrate the skills. This would result in the provision of data to teachers in relation to the skills that they may not gather otherwise, possibly as a consequence of not knowing how the skills are being demonstrated by students in the classroom. The assessment should allow teachers to observe the skills in action and familiarise themselves with them. They should also provide reports that associate observable behaviours with levels of development in the skill. Then teachers can start to make connections between the behaviours they observe in the classroom with the skills. To develop this, data collection needs to continue on a larger scale so that evidence of the behaviours being demonstrated and evidence of how these behaviours link together to form levels of development can be captured. Ideally, these would be built using the template model developed during this project but on a custom-built platform that allows for process data capture including chat logs and automation of scoring.

9.3 Implications for curriculum

Modelling how the general skill development frameworks can be applied in different learning areas

One particular area of interest is to assist teachers in being better able to recognise how the generalisable aspects of critical thinking, creative thinking and collaboration manifest within the specific methodologies, conventions and 'ways of knowing' of their own learning areas. What would it look like to 'identify patterns and make connections' in history-based investigations, English literature analyses or health research? To 'generate ideas' and solutions that have 'novelty' in mathematics, LOTE or design technology, or to 'maintain shared understanding' in group science experiments or drama productions? Making these links explicit will be crucial: if the general capabilities are assumed – but not shown – to operate within learning areas already, and are taught implicitly, students and teachers will not recognise that they are employing them. Conversely, if the skills are taught explicitly but discretely and in isolation from subject-area content or skills, their transfer from a general to a specific context is not assured. In either case, a coherent meta-understanding of what it is that is general about the general capabilities, and unifies these skills across disciplines, will not develop without them being explicitly identified and embedded in the learning areas.

9.4 Implications for pedagogy

Teacher training

Teacher education is another area in which there has been very little advocacy. It is essential that the higher education sector acknowledges the need for review of pre-service education and consider the pedagogical and assessment changes implied by additional learning goals. Professional development is required to train teachers in how the skills are demonstrated in the classroom, how to integrate the assessments in the classroom, how to interpret the assessment reports and use the data, and how to integrate the teaching materials in their classroom.

The popularity of the half-day assessment in general capabilities masterclass at the 2019 ACER Research Conference prompted the delivery of day-long masterclasses for teachers in various state capitals in 2020 (and now delivered online as a four-week course). In these sessions, teacher participants are guided through the process of unpacking ACER's skill definitions and developing a PBL unit, in the mode of the general capabilities assessment, through which to assess these skills. Anecdotally, feedback received from participants and demonstrations of the units of work they have developed, suggests the masterclass has been well-received and that the skill definitions are directly useful to classroom-based assessment. Through the masterclasses, application of the skills development frameworks as well as video samples of students that exemplify student behaviours at various levels of performance are being collected and a subsequent report will outline these findings in more detail.

Teacher strategies and resources

Availability of suitable teaching resources is a major issue for teachers. Materials need to consist of definitions of the skills, classroom-based tasks that teach and elicit the skills, rubrics to score these tasks, and levels of skill development to support them to monitor growth in the skills. To develop this, research needs to identify sufficient data to build evidence-based materials. It is hoped that this project can maintain the engagement of a self-selecting group of teachers who have attended or completed the masterclass and establish a research community with them, wherein strategies, resources and updated forms of assessment can be trialled and tested.

It is acknowledged that the sorts of skills increasingly expected of school graduates and employees can manifest themselves in an enormous range of expressions, contexts and applications that are beyond the scope of a small suite of classroom tasks to definitively assess. Nevertheless, it is important to work to find well-considered and reliable ways teachers can elicit, isolate and measure such nebulous skills. This project has made progress in showing that such skills can be assessed within a framework that also lends itself as a model for teaching the skills. Once this has been achieved, and using correctly-tuned, teacher-friendly assessment methods, it is hoped that schools will continue to propagate the ACER general capabilities project's approach to further embed skills more comprehensively in curriculum, assessment and pedagogy. To be sustainable and a significant contribution to the field, this work needs to be grounded in an evidence-based approach therefore the validation and iterative approach to better understanding the skills continues to be a priority for ACER.

10 REFERENCES

ACARA (n.d.). *General capabilities curriculum*. https://www.australiancurriculum.edu.au/f-10-curriculum/general-capabilities/

Baghaei, N., Mitrovic, A., & Irwin, W. (2007). Supporting collaborative learning and problem solving in a constraint-based CSCL environment for UML class diagrams. *International Journal on Computer-Supported Collaborative Learning*, *2*(2), 159–190.

Care, E., Griffin, P., Scoular, C., Awwal, N., & Zoanetti, N. (2015). Collaborative problem solving tasks. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 85–104). Springer Netherlands.

Care, E., Anderson, K., & Kim, H. (2016). *Visualising the breadth of skills movement across education systems*. https://www.brookings.edu/research/visualizing-the-breadth-of-skills-movement-across-education-systems

Ercikan, K., & Oliveri, M. E. (2016). In search of validity evidence in support of the interpretation and use of assessments of complex constructs: Discussion of research on assessing 21st century skills. *Applied Measurement in Education, 29*(4), 310–318. https://doi.org/10.1080/08957347.2016.1209210

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Erlbaum.

Greiff, S., Wüstenberga, S., Csapo, B., Demetriou, A., Hautamäki, J., Graesser, A. C., & Martin, R. (2014). Domain-general problem solving skills and education in the 21st century. *Educational Research Review, 13*, 74–83.

Heard, J., Scoular, C., Duckworth, D., Ramalingam, D., & Teo, I. (2020). *Critical thinking: Skill development framework*. https://research.acer.edu.au/ar_misc/41/

Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review, 16*(3), 235–266.

Maudsley, G. (1999). Do we all mean the same thing by 'Problem-based learning'? A review of the concepts and a formulation of the ground rules. *Academic Medicine*, *74*(2), 178–185.

Parker, R., & Thomsen, B. S. (2019). *Playful integrated pedagogies that foster children's holistic skills development in the primary school classroom*. LEGO Foundation. https://research.acer.edu.au/learning_processes/22

Ramalingam, D., Anderson, P., Duckworth, D., Scoular, C., & Heard, J. (2020). *Creative thinking: Skill development framework*. https://research.acer.edu.au/ar_misc/40/

Reckase, M. D. (2009). *Multidimensional item response theory.* Springer.

Reeves, T. C., Herrington, J, & Oliver, R. (2002). Patterns of engagement in authentic online learning environments. *Australasian Journal of Educational Technology*, *19*(1), 279–286.

Roseth, V. V., Valerio, A., & Gutierrez, M. B. (2016). *Education, skills, and labor market outcomes: Results from large-scale adult skills surveys in urban areas in 12 countries*. http://documents.worldbank.org/curated/en/397871467995623739/Main-report

Roth, A., Kim, H., & Care, E. (2017, 31 August). *New data on the breadth of skills movement: Over 150 countries identified*. https://www.brookings.edu/blog/education-plus-development/2017/08/31/new-data-on-the-breadth-of-skills-movement-over-150-countries-included/

Scoular, C., Duckworth, D., Heard, J., Ramalingam, D. (2020) *Collaboration: Skill development framework*. Retrieved from https://research.acer.edu.au/ar_misc/42/

Scoular, C. (2019). A design template for transforming games into twenty-first century skills assessments. *Journal of Applied Research in Higher Education*. https://www.emerald.com/insight/content/doi/10.1108/JARHE-02-2018-0018/full/html

Soller, A. (2001). Supporting social interaction in an intelligent collaborative learning system. *International Journal of Artificial Intelligence in Education, 12*, 40–62.

Sternberg, R. J. (1998). *Thinking and problem solving* (Vol. 2). Elsevier.

Webb, N. M., Nemer, K. M., Chizhik, A. W., & Sugrue, B. (1998). Equity issues in collaborative group assessment: Group composition and performance. *American Educational Research Journal, 35*(4), 607–651.

World Economic Forum. (2016). *The future of jobs: Employment, skills and workforce strategy for the fourth industrial revolution.* http://www3.weforum.org/docs/WEF_Future_of_Jobs.pdf

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement*. MESA Press.

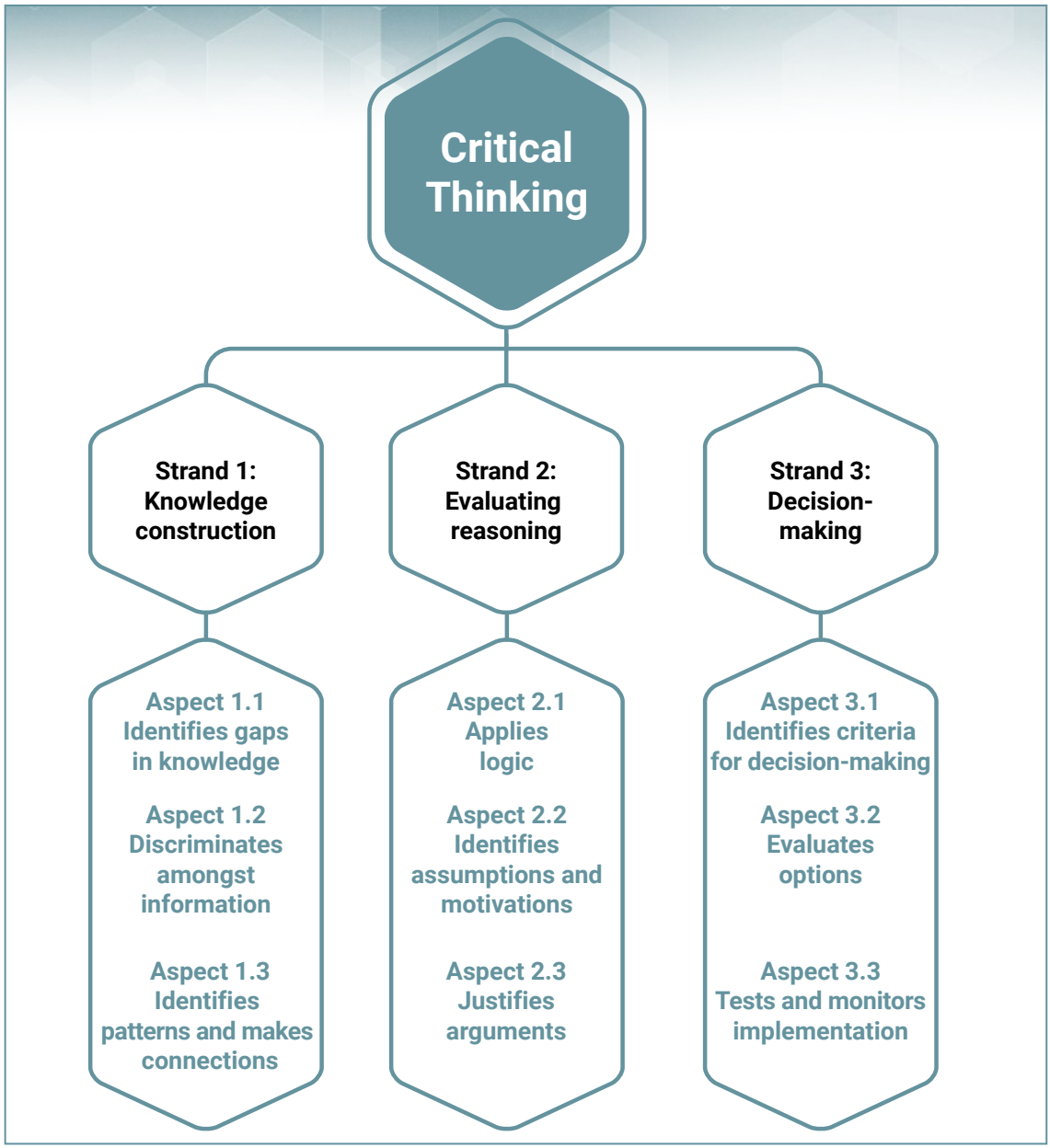
Wright, B. D., & Stone, M. H. (1979). *Best test design*. MESA Press.

Zhang, J. (1998). A distributed representation approach to group problem solving. *Journal of American Society of Information Science, 49*(9), 801–809.

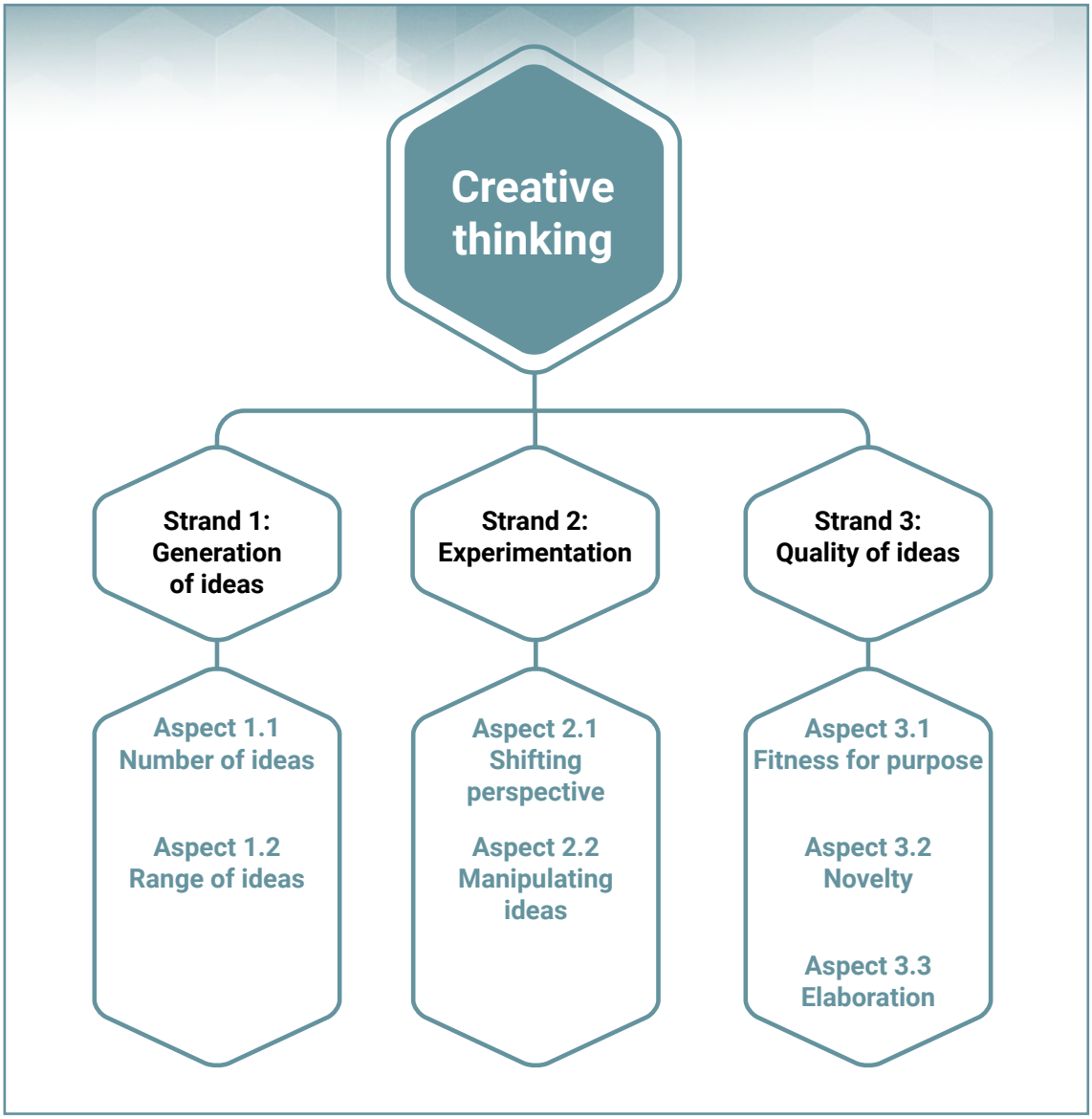
APPENDICES

Appendix A

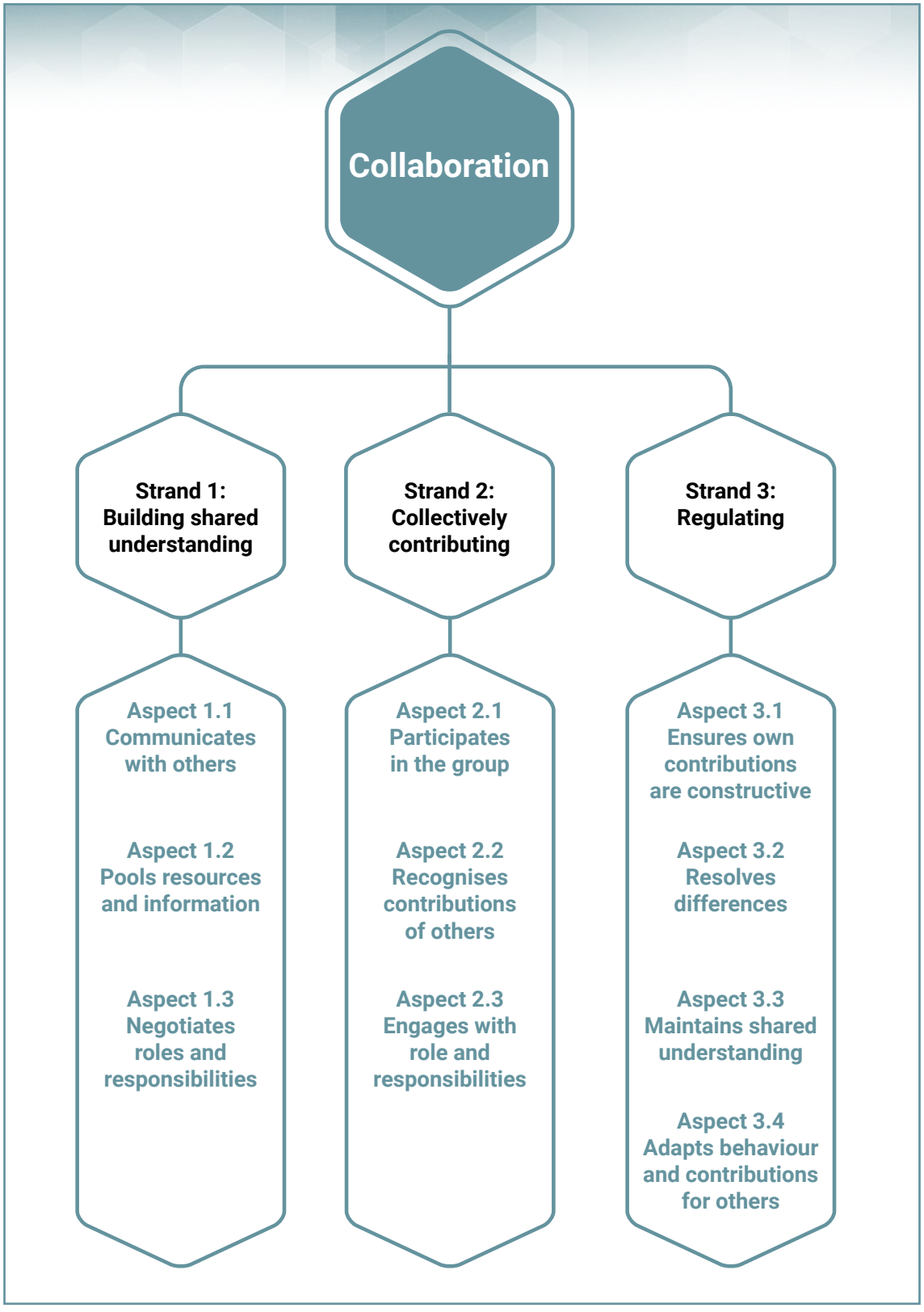
A.1 Skill Development Framework for Critical Thinking from (Heard et al., 2020)



A.2 Skill Development Framework for Creative Thinking from (Ramalingam et al., 2020)



A.3 Skill Development Framework for Collaboration from (Scoular et al., 2020)



Appendix B

Table B.1 Discrepancy rates for marking each item across the two modules

Task	Discrepancy rate (Module 1) (%)	Discrepancy rate (Module 2) (%)
Task 1a	13	11
Task 1b	25	10
Task 1c	15	14
Task 1d	24	19
Task 1e	16	14
Task 1f	12	12
Task 1g	8	N/A
Task 3a*	37	17
Task 3b	58	31
Task 3c	17	17
Task 3d	28	13
Task 3e	26	21
Task 3f	26	20
Task 4bA**	34	37
Task 4bB	33	27
Task 4cA	19	21
Task 4cB	24	21
Task 4cC	19	20
Task 4cD	37	34
Task 4cE	36	34
Task 6a	14	17
Task 6b	7	8
Task 6c	15	17
Task 6d	3	15
Task 6e	11	13
Task 8aA	12	9
Task 8aB	7	10
Task 8aC	6	9
Task 8aD	9	10
Task 8aE	6	12

Task	Discrepancy rate (Module 1) (%)	Discrepancy rate (Module 2) (%)
Task 8aF	7	8
Task 8aG	11	8
Task 8aH	10	7
Task 8aI	7	10
Task 8aJ	21	23
Task 8aK	13	13
Task 8bA	16	14
Task 8bB	20	15
Task 8bC	16	12
Task 8bD	27	19
Task 8bF	30	25
Task 9a	8	18
Task 9b	19	34
Task 9c	21	28
Task 10a	24	18
Task 10b	35	39
Task 10c	35	31

* In some tasks, such as Task 3, scoring of the different parts was somewhat dependent. Hence the discrepancy rates for such tasks are likely to be overinflated.

** For this task only, discrepancy rates were calculated using a subset of the data for that task.